UNIVERSITY OF PENNSYLVANIA

THE SPECIFIC NATURE OF ACHIEVEMENT AND THE PREDICTIVE VALUE OF THE IQ

An Investigation of the Limitations Placed Upon the Predictive Values of Otis IQ's by the Variation in the Achievements of Twelfth Grade Students in Their Different Subjects

HANS COCHRANE GORDON

A THESIS

IN EDUCATION

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

PHILADELPHIA
1931

THE SPECIFIC NATURE OF ACHIEVEMENT AND THE PREDICTIVE VALUE OF THE IQ

An Investigation of the Limitations Placed Upon the Predictive Values of Otis IQ's by the Variation in the Achievements of Twelfth Grade Students in Their Different Subjects

---

A THESIS

IN EDUCATION

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF PENNSYLVANIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

HANS COCHRANE GORDON

PHILADELPHIA
1931

*24017

Copyright 1931
by
Hans Cochrane Gordon

R. UNIV.
BIBLIOTHEEK
LEIDEN

ACKNOWLEDGMENTS

Practically every project in educational research is to some degree a cooperative undertaking. This study is no exception. Without generous access to the Carnegie Program records of the students in the upper twelfth grades of the public schools of Philadelphia, this study would not have been possible. These records were made available through the courtesy of Dr. Philip A. Boyer, Director of the Division of Educational Research and Results of the Philadelphia Public Schools. To him and to the members of the staff of the division, Dr. Harriet M. Barthelmess, Dr. Walker W. Cheyney, Mr. Mark H. Hagman, and Mr. Merle P. Ekas, the writer is indebted for timely suggestions throughout the progress of the work.

For guidance in the formulation of the problem, for very helpful criticism in its half completed stage, and for numerous improvements in the final report, the writer wishes to acknowledge his debt to the Seminar in Secondary Education and to the Group Committee, Dr. E. Duncan Grizzell, Dr. Walter H. Magill, and Dr. LeRoy A. King. As chairman of the Group Committee, Dr. Grizzell has viewed the problem in its larger aspects and has given of his time unsparingly in the determination of the major lines of research. Dr. Magill has devoted many hours to revision of procedures and clarification of the presentation. His numerous and important suggestions have been most welcome. Dr. King has indicated methods of approach which have proved to be quite enlightening. Dr. George G. Chambers, Professor of Mathematics, has contributed generously of his time and advice in checking the appropriateness of the statistical procedures applied.

The patient encouragement of my wife, Elizabeth Hall Gordon, exemplified in unstinted assistance through the more laborious parts of the study, has furnished a stimulus for research which continues our own pursuit of the unknown.

H. C. G.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547328_0012

CONTENTS

PAGE

CHAPTER I

THE PROBLEM 13
I. The Significance of the Problem 13
II. Formal Statement of the Problem 20

CHAPTER II

WHAT OTHER RESEARCH WORKERS HAVE CONTRIBUTED TOWARD THE SOLUTION OF THE PROBLEM 22
I. Relationship of Measures of Mental Ability to Achievement in Specific Subjects 22
II. Predictive Efficiency of the IQ for Achievement of Individuals in Specific Subjects 30
III. The Relative Value of IQ's in Predicting Specific Subject Achievement in Different Parts of the IQ Range 34
IV. The Extent of Variation in Individual Achievement in Different Subjects 34
V. The Ratio of the Average Standard Deviation of the Individual Scores to the Standard Deviation of the Most Complete Group Distribution 36
VI. The Extent to Which Achievement in Specific Subjects Can be Predicted from Average Achievement 37

CHAPTER III

DESCRIPTION OF THE DATA 38
I. The Testing Program 38
II. Validity and Reliability of the Subject Matter Tests 40
III. Reliability of the Average School Marks 46
IV. Validity and Reliability of the Measures of Mental Ability 49

CHAPTER IV

THE MEASURES USED 52
I. The Measure of Specific Subject Achievement: The Subject Criterion 52
II. The Measure of Mental Ability 63
III. The Measure of Relationship 65

CHAPTER V

CORRELATION COEFFICIENTS FOR OTIS IQ'S AND SUBJECT CRITERION SCORES 67

CHAPTER VI

THE PREDICTIVE EFFICIENCY OF THE IQ FOR SPECIFIC SUBJECT ACHIEVEMENT 81

CHAPTER VII

The Relationships Between IQ's and Subject Criterion Scores in Different Parts of the IQ Range 86
I. Correlation Coefficients between IQ's and Subject Criterion Scores for IQ's of 110 or Over and IQ's Lower Than 110 86
II. Criterion Score Means of the Successive Arrays at Different IQ Levels Compared with the Regression Lines for Total Distributions 89

CHAPTER VIII

The Extent of Individual Variation in Specific Subject Achievement and the Limitations Placed Upon the Prediction of Specific Subject Achievement by the Use of Any Index of General Achievement 98
I. Preparing Equated Scales of Specific Subject Achievement 98
II. Profile Charts 101
III. The Correlations between Average and Specific Achievement .. 104
IV. The Efficiencies of Measures of General Achievement in Predicting Specific Achievement 106
V. The Ratio of the Variation of Achievement About Individual Levels to Variation about Group Levels of Achievement 108

CHAPTER IX

Summary and Conclusions 113
I. Critical Study of the Effect of Variation, Reliability, and Validity of Measures Upon the Predictive Values of the IQ's ... 117
II. Possibilities and Limitations of the Use of Otis IQ's for Prediction of Specific Subject Achievement 120
III. The Significance of the Departure of Individual Subject Achievement from Individual Average Achievement 122
Bibliography 123
Appendix 135

TABLES

PAGE

I

Correlation Coefficients between Scores in the Subject-Matter Tests and Average School Marks in the Subjects 42

II

Validity Coefficients for Subject-Matter Tests Used in the Cooperative Program 44

III

Correlation Coefficients betweenScores in Subject-Matter Tests and the Number of Terms Instruction 46

IV

Reliability Coefficients Reported for Tests Used in the Cooperative Program 47

V

Validity Coefficients of the Otis Self-Administering Test of Mental Ability, Higher Examination 50

VI

Transmutation of School Marks in English to Linear Scores 53

VII

Linear Score Values of Centiles 54

VIII

Transmutation of Ten Columbia Research English Test Scores to Linear Scores 56

IX

Raw Correlation Coefficients between Otis IQ's and Subject Criterion Scores Ranked According to Magnitude and Compared with the Same Coefficients Corrected for Differences in the Variation of IQ's... 68

X

Criterion Score and IQ Means and Standard Deviations for Each Subject.. 71

XI

Correlation Ratios (Etas) between Otis IQ's and Subject Criterion Scores Compared with Pearson Product Moment Correlation Coefficients (r's) 73

XII

Correlation Coefficients between (1) Scores in Otis Test of Mental Ability, Higher Examination and Subject Achievement Test Scores, (2) Otis IQ's and Average School Marks and (3) Otis IQ's and Subject Criterion Scores 74

XIII

Correlation Coefficients between Measures of Mental Ability and Achievement in English, Ranked According to Magnitude 76

XIV

Correlation Coefficients between Measures of Mental Ability and Achievement in Social Studies, Ranked According to Magnitude 77

XV

Correlation Coefficients between Measures of Mental Ability and Achievement in Mathematics, Ranked According to Magnitude 78

XVI

Correlation Coefficients between Measures of Mental Ability and Achievement in Physical Sciences, Ranked According to Magnitude 78

XVII

Correlation Coefficients between Measures of Mental Ability and Achievement in Foreign Languages, Ranked According to Magnitude 79

XVIII

Per Cents Reduction in the Errors of Predicting Specific Subject Achievements Using Otis IQ's Rather than Pure Chance Placement within the Actual Subject Groups 82

XIX

Per Cents Reduction in the Errors of Predicting Specific Subject Achievements Using Otis IQ's Rather than Pure Chance Placement within Hypothetical Subject Groups of the Same IQ Variation as That of the Total Group 83

XX

Correlation Coefficients Between IQ's and Subject Criterion Scores for Upper and Lower IQ Groups: Raw and IQ Variation Corrected Coefficients 88

XXI

Correlation Distribution for IQ's and Criterion Scores in English 90

XXII

Numbers of Units Added to (+) or Subtracted from (−) the Criterion Scores in Each Subject to Produce Comparable Scales at the Same Levels..... 100

XXIII

Raw Correlation Coefficients between the Mean Criterion Score for Each Individual and Each Separate Criterion Score and the Same Coefficients Corrected to the Variation of the Mean Scores of the Total Group of 183 Students..... 105

XXIV

Per Cents Reduction in the Errors of Predicting Specific Subject Achievements Using Mean Criterion Scores Rather than Chance Placement. (Hypothetical Subject Groups with the Same Variation of Mean Criterion Scores)..... 106

XXV

Ratios of the Standard Errors of Estimating Criterion Scores from Individual Mean Criterion Scores to the Standard Deviations of the Complete Distributions for 183 Students..... 109

XXVI

Standard Deviations of the Separate Subject Criterion Scores About Individual Mean Criterion Scores, Expressed as Per Cents of the Standard Deviations of the Complete Criterion Score Distributions.. 110

FIGURES

1
Transmutation of School Marks in English to Linear Score Units, Showing the Frequency Surfaces on the Normal Frequency Distribution for Each School Mark and Its Transmuted Linear Score..... 55

2
Interpolation of Linear Score Values for Raw Scores in Columbia Research English Test..... 57

3
Distribution of Average School Marks in English..... 59

4
Distribution of Linear Score Values of Average School Marks in English... 59

5
Distribution of Raw Scores in Columbia Research English Test..... 59

6
Distribution of Linear Score Values of Columbia Research English Test Scores..... 60

7
Superimposed Distributions of Linear Score Values of Average School Marks and Test Scores in English..... 60

8
Distribution of Criterion Scores in English..... 61

9
Comparison of the Means of the English Criterion Score Arrays with the Regression of English Criterion Upon IQ..... 93

10
Comparison of the Means of the Algebra Criterion Score Arrays with the Regression of Algebra Criterion Upon IQ..... 94

11
Comparison of the Means of the American History Criterion Score Arrays with the Regression of American History Criterion Upon IQ..... 95

12
Comparison of the Means of the Physics Criterion Score Arrays with the Regression of Physics Criterion Upon IQ..... 96

13
Comparison of the Means of the Latin Criterion Score Arrays with the Regression of Latin Criterion Upon IQ..... 97

14
Individual Profile Chart Showing Widely Variable Criterion Scores..... 102

15
Individual Profile Chart Showing Closely Sim lar Criterion Scores..... 103

16
Probable Errors of Estimate of Criterion Scores Using: 1. The Average of Five Criterion Scores, 2. Otis IQ..... 107

17
Superimposed Distributions of 915 Criterion Scores for 183 Individuals Based Upon Intervals of 5 Criterion Score Units Measured from: 1. The Individual Means, 2. The Group Means..... 111

CHAPTER I

I

The Problem

I. *The Significance of the Problem*

Knowledge of the relationships existing between measures derived from general mental ability tests and measures of school achievement is fundamental to the utilization of such tests for practical school purposes. A "mental ability test" or an "intelligence test" may be defined for the purposes of this study as follows: a test of performance in certain tasks so selected that the score is a measure of those non-specialized mental characteristics of individuals by which they learn to perform an increasingly wider range of abstract mental activities. Freeman defines a mental test as follows:

> "Mental tests are instruments for the measurement of relative mental capacity, either general or special."[1]

The definition used in this study limits the type of test which can be classified as a mental ability test more rigidly to instruments for measuring general attributes rather than the specific attributes which prognostic tests for specific subjects attempt to measure. The definition used in this study also excludes tests of achievement, either in specific subjects, as the Columbia Research English Test, or in groups of subjects, as the Stanford Achievement Tests. Tests which are illustrative of mental ability tests in the sense of the definition used in this study are as follows:

Otis Self-Administering Tests of Mental Ability
Terman Group Test of Mental Ability
Binet Simon Test (Stanford and Herring Revisions)
Haggerty Intelligence Test, Delta 1 and 2
Illinois Examination I and II
Philadelphia Mental Ability Test
Thorndike Intelligence Examination for High School Graduates

[1] Freeman, F. N.: *Mental Tests*, 20, 21.

Whether scores in so-called mental ability tests and measures such as the IQ derived from these scores are indices of intellectual capacity may be questioned. However, one of the values of such scores and their derived measures is due to their known ability to predict for some pupils actual performance in certain school tasks. Without limiting the utility of measures of general mental ability so-called, it may be said that what they purport to measure is probably too generalized a quantity for measurement at the present time. Kelley says in speaking of the classification of pupils:

> "Another unwarranted category is that of general intelligence when used in the sense that those high in this trait are possessed of more ability in all mental traits than those low in it."[2]

However, recognition of the fact that general mental ability now may be incapable of measurement as such only serves to center attention upon the immediate functions which can be served by mental ability tests. These functions generally involve the prediction of certain types of school achievement.

If school success can be predicted by mental ability measures, then these should be adaptable tools in the hands of the educator. For, the same power of prediction may be used by the teacher or school administrator in the educational guidance of pupils, in assigning pupils to groups based upon probably homogeneity of performance, and in setting as standards for individuals or groups achievement levels comparable with the average achievement of pupils with similar mental ability test scores. Educational research workers also may evaluate the actual achievements of pupils in terms of their predicted achievements and to this extent evaluate instruction. Experimental techniques may be set up to evaluate different educational procedures by the selection of control and experimental groups equated to the extent that predicted achievements are similar.

In each case where measures of mental ability are used for predictive purposes, it is essential that information be at hand

---

[2] Kelley, T. L.: *Interpretation of Educational Measurements, 116.*

to indicate the degree of relationship between the achievement to be predicted and the mental ability measure used. Any prediction involving mental and social measurements is subject to a very considerable error. Accordingly, prediction must be only in terms of probabilities. With this limitation, evidence may be cited to show that the utilization of measures of mental ability to predict a single average for achievement in all academic subjects is sanctioned in some measure both by current practice and by the findings of investigators relative to the validity of intelligence tests. Since the War, when the fact was disclosed that various occupational and educational groups showed decided differences in range and average of mental ability scores, there has been a steadily mounting accumulation of such evidence. Of 194 published articles listed by Odell[3] in a bibliography dealing with the subject, 58 express opinions or cite data relative to the value that intelligence tests have in the prediction of average achievement in school. The tests may be used alone or in conjunction with other predictive indices. Fifty of these studies give a favorable estimate of the value of intelligence tests in predicting ability in school.

The use of scores in so-called intelligence tests as indices of ability is rather widespread. An investigation by Dvorak[4] found that, of 86 schools grouping pupils homogeneously, 45 used mental ability tests, and 27 used other educational tests.

Another investigator[5] analyzed the returns from 50 junior and senior high schools in different parts of the United States. Slightly more than half of this number classified students on the basis of intelligence test scores and teachers' estimates. Two-fifths used intelligence test scores and school marks.

An investigation[6] in June 1930 by the Division of Educational

---

[3] Odell, C. W.: "An Annotated Bibliography Dealing with the Classification of Pupils to Provide for Individual Differences." *University of Illinois Bulletin*, 21: 12: 5–32. November 19, 1922.

[4] Dvorak, A.: "Recognition of Individual Differences in the Junior High School." *School Review* 30: 679–685. November 1922.

[5] Omans, A. C.: "Provision for Ability Grouping in Junior and Senior High Schools." *American School Board Journal*, 65: 55–58. October 1922.

[6] Philadelphia Board of Public Education, Division of Educational Research. Bulletin 175, *Grouping by Ability and Individualization of Instruction.*

Research of the Philadelphia public schools shows that IQ's are used more frequently than any other index of probable school performance in the junior high schools of that city. All of the 20 junior high schools report the use of mental ability tests in sectioning pupils according to ability. In the senior high schools 7 of the 13 schools use mental ability tests in sectioning by ability, although but few schools report any grouping except that entirely within subject fields.

In the colleges and universities, Toops[7] found that the use of "intelligence" tests was increasing. One hundred and ten questionnaires were sent to colleges, and one hundred and ten replies were returned. Yearly growth in the number of colleges using intelligence tests is shown by successive increases for these colleges from 1919 to 1926 as follows:[8]

YEAR	NUMBER OF COLLEGES USING INTELLIGENCE TESTS
1919–1920	8
1920–1921	30
1921–1922	39
1922–1923	48
1924–1925	59
1925–1926	66

The validity of measures of mental ability as indices of average success in school is indicated by a number of writers. Thorndike[9] estimated the correlation coefficient between intelligence and success in elementary schools to be .80. His estimated coefficient for high schools is .60, and he adds,

"but if all were forced to try to do this advanced work, the correlation coefficient would be .80 or more."[10]

Cheyney[11] finds a correlation coefficient of .666 between

[7] Toops, H. A.: "The Use of Intelligence Tests by Universities," *School and Society*, 24: 603. July 17, 1926.
[8] *Loc. cit.*
[9] Thorndike, E. L.: "Intelligence and Its Uses." *Harper's Magazine*, 140: 227–235. 1919.
[10] Thorndike: *op. cit.*, 233.
[11] Cheyney, W. W.: *Rate of Progress and the IQ*, 29.

Otis IQ's and rate of progress determined by the time required to complete the first eight grades. Banker[12] reports coefficients of correlation for eight half grades from four to eight for a total of 331 pupils between a student's ability index and Otis IQ's. The students' ability index is derived from the students' average marks in the grade. The average coefficient reported was .825.

Miller[13] reports correlation coefficients of .522, .586, .60, between high school freshman marks and scores in (1) the Miller Mental Ability Test, (2) the Terman Test, and (3) the average of the Miller, Haggerty Delta 2, Terman, Army Alpha, Trabue's Mentimeters, and the Otis tests respectively. However, he believes that the true relationship is higher because he says,

> "Higher correlations between mental tests and school marks than are now obtained can not be expected until marks are based more exclusively on achievement."[14]

Flemming, another investigator,[15] obtains a correlation coefficient of .63 between achievement criteria and intelligence test scores in the senior high school. In the junior high school, coefficients of .60 and .52 are reported between achievement criteria and scores in the Terman and Otis tests respectively.

As part of the current study, the author found the correlation coefficient of .617 between Otis IQ's and the average of criterion scores[16] in 5 subjects for 183 upper twelfth grade students in Philadelphia's senior high schools. The IQ variation of these 183 students is represented by a standard devia-

---

[12] Banker, H. J.: "Correlation Studies of the Students' Ability Index." *Journal of Educational Research*, 20: 31–37. June, 1929.

[13] Miller, W. S.: "The Administrative Use of Intelligence Tests in the High School." *National Society for the Study of Education 21st Year Book:* 189–222. 1922.

[14] Miller: *op. cit.*, 211.

[15] Flemming, C. W.: *A Detailed Analysis of Achievement in the High School; Comparative Significance of Certain Mental, Physical, and Character Traits for Success.*

[16] The term "criterion scores" is used in this study in the meaning of "measures of achievement." The criterion score here described consists of a single measure which is a composite of two equally weighted elements: (1) average school mark in the subject, and (2) a standardized test score. The method of deriving criterion scores is described on pp. 52–63.

tion of 9.0, while that for the entire group of 2209 students in the upper 12th grade was 9.6. Therefore, a larger correlation coefficient may be accepted as more nearly correct if it were possible to determine the coefficient for the entire group.[17]

Probably because of the limitation in variation, the correlation coefficients between intelligence test results and average college grades are lower than similar results in the high or elementary schools. A recent investigation[18] reports a coefficient of .45 between scores in the Thurstone Psychological Examination and average scholastic rating of students at Colorado State Teachers' College. Toops[19] summarizes 43 determinations of correlation coefficients in different colleges between intelligence test results and average scholarship grades for freshmen in a distribution as follows:

CORRELATION INTERVAL	NUMBER OF COLLEGES REPORTING
.60 or over	6
.56–.59	0
.52–.55	7
.48–.51	6
.44–.47	7
.40–.43	8
.39 or under	9
Total	43
Median	.46

It would seem evident from the foregoing brief summary of research that measures of mental ability are not only widely used but that they have some value in predicting general, or average standing in elementary, secondary, and higher schools. However, in secondary schools, prediction of general achievement is seldom of as much value as the prediction of achieve-

[17] This correction is discussed on pp. 69–70.

[18] Condit, P. M.: "The Prediction of Scholastic Success by Means of Classification Examinations." *Journal of Educational Research*, 19: 333. May, 1929.

[19] Toops, H. A.: "The Status of University Intelligence Tests in 1923–24." *Journal of Educational Psychology*, 17: 23–26, 110–124. January and February, 1926.

ment in the specific subjects of the curriculum. An example of the application of a group intelligence test to the prediction of specific subject achievement is given in P. M. Symonds' Ability Standards[20] where a pupil of a given chronological age and score in the Terman Group Test of Mental Ability is assigned scores of 9 high school achievement tests as "Ability Standards." As a matter of fact, when intelligence tests are used as bases, either wholly or in part, of general ability groups in a departmentally organized school, one of the assumptions is that an intelligence test score is an index of ability in algebra during the first period of the day, Latin during the second period, science during the third and fourth periods, etc. In other words, as indices for grouping in the high school, intelligence test scores or derived measures mean little unless they are able to predict specific achievement in the various subjects. Insofar as educational guidance is concerned with pupil election of specific subjects, the use of intelligence test scores as forecasts of probable achievement is also subject to the limitations of their value as indices of specific subject achievement. Similar observations may be made for the use of measures of mental ability in evaluation of instruction. In brief, most uses of mental ability tests in high schools depend upon their power to predict specific rather than general achievement.

The fact that achievement in one subject or even average achievement in a number of subjects does not forecast achievement in another subject is common experience. As an illustration, the result of a grouping experiment in one[21] of Philadelphia's high schools may be cited. A small group of students was selected in the twelfth grade to do work of more intensive character than the regular course. The work planned for these students was somewhat in the nature of an "honors course". The basis of selection was the average of the recorded school marks of the students. While the plan was successful in some ways, the most obvious difficulty was that even in this selected group, the demonstration of superior ability in languages was no guarantee of similar performance in mathematics. The

[20] Symonds, P. M.: *Ability Standards.*
[21] Unpublished study in the Philadelphia High School for Girls.

variation in specific subject achievement was but little reduced by a grouping dependent upon an average mark in a number of subjects as a measure of general ability.

Not only is it desirable to know for each specific subject the predictive power of scores in so-called intelligence tests or measures derived from these scores, but also whether or not these ratings are more valuable for predictive purposes in some parts of their range than in other parts. The findings of this study seem to indicate that IQ's are more satisfactory indices of specific achievement in certain parts of their range than they are in other parts. Thus IQ's over 130 seem to indicate more accurately the distribution position of criterion scores in American history and English than those below 130.[22]

Summarizing, it may be said that the practice of using measures of mental ability in forming a predictive estimate of the achievement of pupils is widespread and increasing. In the secondary school, these measures of mental ability predict average achievement in academic fields to an extent indicated approximately by a correlation coefficient of .60. However, the use of mental ability measures in the secondary school involves more often the prediction of achievement in specific subject fields. Investigation of this type of prediction involves research into the departure of specific achievement from the individual average achievement for all subjects.

II. *Formal Statement of the Problem*

The major problem of this study is to determine the extent to which achievement in specific academic subjects in the high school can be predicted by IQ's[23] derived from scores in the Otis Self-Administering Test of Mental Ability, Higher Examination, Form A. The problem is resolved into six different phases, as follows:

---

[22] See pp. 93–97.

[23] IQ's are used in the above determinations because they are more generally applied than other measures of mental ability. The findings are equally true for mental ages, raw test scores, and I. B.'s because, at the age of 18 years, or the level of graduation from senior high school, IQ's are no more than transmuted raw scores of the test. A similar fact is true of mental ages and I. B.'s.

1. What are the relationships expressed as correlation coefficients between the Otis IQ's and achievements in English, French, German, Latin, Spanish, chemistry, physics, algebra, plane geometry, solid geometry, trigonometry, American history, ancient history, European history, economics, and civics determined for students in the graduating classes of the Philadelphia senior high schools in May, 1928?

2. What is the value of an Otis IQ in the prediction of individual achievement in each of the above subjects?

3. What differences in predictive value exist for IQ's in different parts of the IQ range? Is subject achievement predicted better by high, middle or low IQ's?

4. What is the variation of specific subject achievement from individual average achievement?

5. How does the average standard deviation of individual scores compare with the standard deviation of the most complete group distribution?

6. To what extent can achievement in specific subjects be predicted from individual average achievement?

CHAPTER II

What Other Research Workers Have Contributed Toward the Solution of the Problem

While little has been done to determine the specificity of achievement or the degree to which specific subject achievements of individuals differ from their average achievements at the upper twelfth grade level, considerable data bearing upon the problem have been collected by research workers. The immediate problem is concerned with secondary schools, yet some light is thrown upon the question by the following data from elementary and higher education. Investigations in psychology also have contributed in some measure to the solution of the problem.

I. *Relationship of Measures of Mental Ability to Achievement in Specific Subjects*

Probably more data have been assembled by research workers to determine the relationships between measures of mental ability and achievement in specific subjects than for any other phase of the problem. Nevertheless, exact and precise statements of these relationships based on the assembled data are impossible. The reason for this can be attributed both to the lack of general agreement as to what constitutes mental ability and specific subject achievement and to somewhat unreliable methods of measuring them. A further cause for variation in the results of different investigations is the differences in groups tested. These facts, together with some generalizations of a rather broad nature, will be evident in the summary of reported research immediately following.

Kefauver[1] in a determination of the validity of different bases for ability grouping in the 7th grade of the junior high schools of Fresno, California, calculated zero order correlation coefficients

---

[1] Kefauver, G. N.: "The Validity of Bases for Forming Ability Groups." *Teachers' College Record*, 31: 2: 99–114. November, 1929.

between measures of success in English, arithmetic, and social science and scores in the Multimental Test and the IQ's so derived. These correlation coefficients are given below:[2]

FACTOR	CORRELATED WITH:		
	ENGLISH	ARITHMETIC	SOCIAL SCIENCE
Multimental Score	.37	.42	.42
Multimental IQ	.51	.42	.45

Symonds[3] in "Ability Standards" presents a table of correlation coefficients for scores in the Terman Group Test of Mental Ability and scores in certain achievement tests. From the table median correlation coefficients are determined and reported below:

SCORES IN TERMAN GROUP TEST OF MENTAL ABILITY CORRELATED WITH:	MEDIAN CORRELATION COEFFICIENT
Haggerty Reading Test Sigma III	.77
Hotz Algebra Test—Equations	.27
Hotz Algebra Test—Problems	.45
White Latin Test	.46
Schorling-Sanford Geometry Test	.57
Van Wagenen American History Test	.55
Ruch-Popenoe General Science Test	.60
American Council French Test	.30
Ruch-Crossman Biology Test	.52
Powers Chemistry Test	.50

Allen[4] determined criterion scores[5] for 364 boys electing Latin in the Boys' High School of Brooklyn by obtaining the sum of the weighted scores in eleven specially prepared Latin tests. The correlation coefficient[6] between scores in the Terman Group Test of Mental Ability and the Latin criterion was .439. In a somewhat similar study, Clem[7] reports correlation coefficients

[2] Kefauver, *op. cit.*, 113.
[3] Symonds, *op. cit.*, 16.
[4] Allen, W. S.: *A Study in Latin Prognosis.*
[5] The expression "criterion score" as it is used throughout this study means a numerical measurement of the most inclusive sort obtainable to be used as a criterion of achievement. The criterion score is usually a composite.
[6] Allen, *op. cit.*, 13.
[7] Clem, O. M.: "Latin Prognosis." *Journal of Educational Psychology*, 16: 160–169. March, 1925.

of .36, .48, and .61 for IQ's and a Latin criterion. Jordan[8] determined the correlation coefficient for 53 students between teachers' marks in Latin and their scores in the Terman Group Test of Mental Ability to be .53. For 28 French students and 41 Spanish students similar coefficients were .45 and .60 respectively. School marks in foreign languages for 104 pupils correlated with the Terman scores to the extent of .51.

The Modern Foreign Language Study summarizes a number of investigations of the relationship of measures of intelligence to achievement in foreign languages. A study[9] by Florence Hickman at Salt Lake City shows lower coefficients than those reported above: The relationship between the IQ and achievement in Spanish is expressed in coefficients of .19 and .31 for American Council Alpha Spanish Test, Part I, and teachers' marks in Spanish respectively. In French, median coefficients between IQ's and scores in the three parts of the American Council Alpha French Test after differing periods of instruction were determined by the same investigator. Median coefficients are as follows:

PART OF AMERICAN COUNCIL ALPHA FRENCH TEST	NUMBER OF SEMESTERS INSTRUCTION	MEDIAN CORRELATION COEFFICIENT
Vocabulary	I	.25
Vocabulary	IV	.33
Grammar	IV	.30
Grammar	VI	.29
Silent Reading	IV	.42
Silent Reading	V	.41

Similar data are reported[10] for the Cleveland study. The measures of achievement in French were teachers' marks and percentile ranks in the American Council Alpha French Test. IQ's were correlated with these with the following results for the ends of the 7th, 9th, and 11th years respectively:

---

[8] Jordan, J. N.: "Prognosis in Foreign Languages in Secondary Schools." *School Review*, 33: 541–546. September, 1925.

[9] Reported in Henmon, V. A. C., and Others.: *Prognosis Tests in the Modern Foreign Languages*, 6–7.

[10] Henmon, *op. cit.*, 8.

YEAR	CORRELATION BETWEEN IQ AND MARKS IN FRENCH	CORRELATION BETWEEN IQ AND PERCENTILE RANKS[11] IN AMERICAN COUNCIL ALPHA FRENCH TEST
7	.24	.23
9	.36	.24
11	.19	.30

These low correlation coefficients are evidence of the fact that IQ's did not show close relationship to achievement in French. Further evidence is found in the fact that the median IQ of the failures was 109. The Montclair[12] study furnished similar evidence although the coefficients were somewhat larger, as follows:

TERM OF INSTRUCTION	CORRELATION BETWEEN IQ AND MARKS IN FRENCH	CORRELATION BETWEEN IQ AND PERCENTILE RANKS[11] IN AMERICAN COUNCIL ALPHA FRENCH TEST
1	.38	.41
3	.39	.51
5	.33	.44

Other coefficients reported by Henmon[13] were .38 and .41 for Terman scores and criterion scores in French and Spanish respectively. Another determination[14] of the correlation between scores in the Terman Intelligence Test and a French criterion, consisting of teachers' grades and scores in the American Council Alpha French Test, yielded a coefficient of .424. In the Modern Foreign Language Study the following summarizing statement is made:

"The cumulative evidence from these various studies is that the correlation between intelligence test scores and success in the modern foreign languages, whether estimated by teachers' marks or objective tests lies somewhere between .20 and .60, more often between .30 and .40. While this indicates a positive relationship and that an intelligence test score would be a useful

[11] Linear correlation coefficients between IQ's and percentile ranks are likely to be lower than if raw scores are used to replace the percentile ranks.
[12] Henmon, *op. cit.*, 9.
[13] *Ibid.*
[14] Henmon, *op. cit.*, 176.

member in a team of tests to predict accomplishment, it demonstrates that the intelligence test score cannot be used alone successfully."[15]

In four high school subjects other than foreign languages correlation coefficients are reported by Jordan[16] for four separate intelligence tests. Scores in the tests were correlated with school marks in the subject. The correlation coefficients are reported below:

INTELLIGENCE TEST	COEFFICIENT OF CORRELATION FOR:			
	ENGLISH	MATHEMATICS	GENERAL SCIENCE	HISTORY
Otis	.412	.281	.407	.340
Army Alpha	.391	.371	.480	.285
Miller	.428	.342	.466	.147
Terman	.424	.335	.508	.356

Correlation coefficients were determined by Flemming[17] for French, English, mathematics, and history marks with scores in the Terman, Miller, and Otis tests:

INTELLIGENCE TEST	COEFFICIENT OF CORRELATION WITH MARKS IN:			
	FRENCH	ENGLISH	MATHEMATICS	HISTORY
Terman	.329	...	...	.26
Miller	.331	...	...	...
Otis	...	.26	.32	...

Coefficients have been determined for a much larger number of secondary school subjects by Violet Fuller.[18] IQ's for 800 pupils at the State Preparatory School of Boulder, Colorado, were correlated with marks in twenty-three different courses. The obtained coefficients are indicated below:

---

[15] Henmon, *op. cit.*, 12.

[16] Jordan, A. M.: "Correlations of Four Intelligence Tests with Grades." *Journal of Educational Psychology*, 13: 419–429. October, 1922.

[17] Flemming, C. W.: *A Detailed Analysis of Achievement in the High School*, 124.

[18] Fuller, Violet: *The Relation of Intelligence to the Passing of Secondary School Subjects.*

NAME OF THE COURSE	CORRELATION	NAME OF THE COURSE	CORRELATION
Algebra 1	.40	English 1	.51
" 2	.05	" 2	.34
" 3	.05	" 3	.31
Plane Geometry	.45	" 4	.30
Solid "	.36	Latin 1	.45
Trigonometry	.23	" 2	.24
Commerical Arithmetic	.13	" 3	.30
Ancient History	.32	" 4	.27
Modern History	.23	General Science	.62
American History	.16	Biology	.48
Civics	.11	Physics	.27
		Chemistry	.38

One other study in the secondary field has been completed by Rector.[19] IQ's determined from Army Alpha scores for 10th grade students were compared with marks in various subjects. The numbers of students varied from 431 for English to 49 for shorthand. The resulting correlation coefficients are reported below:

SUBJECT	CORRELATION	SUBJECT	CORRELATION
English	.25	Drawing	.097
Mathematics	.28	Arithmetic	.33
Science	.28	Typing	.21
Spanish	.15	Bookkeeping	.22
French	.092	Shop	.13
Latin	.38	Shorthand	.072
Home Economics	.075	History	.33
Music	.29		

An outstanding piece of research in the elementary school field was completed by Gates.[20] Using scores for a composite of 15 intelligence tests, correlation coefficients were obtained with criterion scores in reading, spelling, and arithmetic. The coefficients are somewhat larger than most of those reported for secondary school studies:

---

[19] Rector, W. G.: "A Study in the Prediction of High School Success." *Journal of Educational Psychology*, 16: 28–38. January, 1925.

[20] Gates, A. I.: "The Correlations of Achievement in School Subjects with Intelligence Tests and Other Variables." *Journal of Educational Psychology*, 13: 139. March, 1922.

Reading Comprehension	.72
Reading Rate	.68
Spelling	.56
Arithmetic	.30

Studies by Buckingham, Odell, and Stoddard in the field of higher education indicate that the correlation between intelligence scores and subject achievements are of magnitudes similar to those found by investigators in the secondary field. Thus Buckingham,[21] in an investigation of mathematical ability, reports correlation coefficients for achievements in algebra, geometry, and calculus with scores in the Terman, Chicago, Army Alpha tests, and composites:

Terman and Mark in Algebra	.25
Chicago and Mark in Algebra	.26
Composite Terman and Chicago with Mark in Algebra	.19
Army Alpha and Mark in Algebra	.38 and .23
Army Alpha and Mark in Geometry	.40
Army Alpha and Mark in Calculus	.10

Although most coefficients found by Buckingham were lower than .40, yet he estimates that .40 approximates the true relationship between mathematical ability and intellectual capacity. In summarizing the study, this statement is made:

"There seems to be much in the success of students in mathematics—and in their failure as well—to countenance the belief that mathematical ability is in no small degree sui generis."[22]

Odell's[23] research into the relationship between IQ's and success in foreign languages resulted in the following coefficients:

IQ and Achievement in French	.32
IQ and Achievement in German	.27
IQ and Achievement in Latin	.39
IQ and Achievement in Spanish	.27

[21] Buckingham, B. R.: "Mathematical Ability as Related to General Intelligence." *School Science and Mathematics*, 21: 205–215. March, 1921.
[22] Buckingham, *op. cit.*, 215.
[23] Odell, C. W.: *Predicting the Scholastic Success of College Freshmen.*

As a part of Stoddard's[24] study of the predictive value of scores in the Thorndike Intelligence Examination, correlation coefficients with college marks were determined as follows:

THORNDIKE EXAMINATION SCORES CORRELATED WITH COLLEGE MARK IN:	CORRELATION
Chemistry	.270 and .407
English	.418 and .359
French	.251 and .179
Mathematics	.229 and .173

While not by any means exhaustive of the literature in the subject, the above summary will indicate that considerable data have been collected which bear on the relationship between measures of intelligence and specific subject achievement. Undoubtedly, some of the differences in the reported correlation coefficients are due to unreliability of the correlation coefficients. Some of the results were based upon an insufficient number of cases for reliable determinations and it is probable that the reliabilities of the subject achievement and the mental ability measures differed enough to affect correlation coefficients. Differences in the score variation of the groups measured probably are responsible for some of the scattering of the numerical values of the correlation coefficients. However, certain facts seem evident from an examination of the data presented in these researches:

1. All correlation coefficients between measures of mental ability and specific subject success are positive.

2. Almost all coefficients for specific subjects are lower than the correlation coefficients determined between measures of mental ability and measures of average success in school.

3. The coefficients determined for any one subject show a range so wide that the average coefficient as an estimate of the existing relationships is likely to be in error from at least 10 to 20 per cent.

4. Reading, English, and general science usually show closer relationship to measures of mental ability than do other sub-

[24] Stoddard, G. D.: *Iowa Placement Examinations*, 75–78.

jects for which correlation coefficients were determined under similar conditions.

II. *Predictive Efficiency of the IQ for Achievement of Individuals in Specific Subjects*

From the standpoint of this study, the most significant contributions of other research workers toward the solution of the problem of predictive efficiency have been in establishing units and methods of measurement. A number of different methods have been used. The values of these methods are compared below:

1. The Pearson Product-Moment Coefficient of Correlation.

Although the most commonly used measure of the relationships between mental ability and achievements in specific subjects is the Pearson product-moment coefficient of correlation, this measure is not the most satisfactory for the purpose of expressing predictive power. For example, raw correlation coefficients are sometimes misinterpreted as per cents of perfect prediction. However, some investigators have endeavored to interpret correlation coefficients in more readily intelligible terms.

2. Per Cent of the Total Number of Cases Correctly Sectioned.

One method of interpretation is to divide the group into sections and to estimate the per cent of the entire group placed in the same section by both the measures which have been correlated. In the use of this method, a group may be divided into a number of sections according to the measure of mental ability. The distributions of subject achievement scores of the mental ability sections thus segregated are studied to determine the spread of achievement scores in each mental ability group. The entire group may be re-sectioned on the basis of subject achievement scores. Where the achievement scores and the measures of mental ability agree to the extent that pupils are placed in the same group (as upper, middle, or lower) whether segregated by the mental ability index or the subject criterion, those pupils are said to be correctly placed. Estimates may be made of the per cents correctly placed or displaced by one group, two groups, etc.

As illustrative of this form of predictive expression the following statement may be cited:

"About 16 per cent of the students received a rating which fell in the same decile as their intelligence score decile; about 38 per cent exceeded expectations and about 48 per cent fell below the expected decile of scholarship. The data reveal the fact that 42 per cent of the entire group either fell within the same deciles in both intelligence and scholarship or were displaced no more than one decile either way."[25]

Brooks[26] reports results of his measurement of the value of different bases for ability grouping similarly. Thus 63.4 per cent were correctly sectioned by the Terman and 64.5 per cent by the Stanford-Binet IQ's.

This method of determining the per cents of the total group in correctly placed sections may be valuable for practical administration. However, the method has little value as an accurate measure of predictive efficiency. The per cent of the total number of cases rightly placed is an ambiguous[27] index of test battery efficiency because it varies with the coarseness of the scale. Thus a coefficient of correlation of .60 will place correctly 20 per cent within one-half point of their true aptitude on a 10-point scale, 38 per cent on a 5-point scale, and 59 per cent on a 3-point scale. A correlation coefficient of .00 places correctly 16, 31, and 50 per cent on 10-, 5-, 3-point scales respectively.

3. The Ranges of Achievement at Successive Placement Levels.

Another method of indicating the value of an index for predictive purposes is to describe the ranges of achievement scores or IQ's at successive levels. Thus Henmon[28] reports that pupils receiving "low marks" in French had IQ's ranging from 81 to 131 with a median at 109.

---

[25] Condit, P. M.: "The Prediction of Scholastic Success by Means of Classification Examinations." *Journal of Educational Research*, 19: 331–335. May, 1929.

[26] Brooks, F. D.: "Sectioning Junior High School Pupils by Tests and School Marks." *Journal of Educational Research*, 12: 359–369. December, 1925.

[27] Hull, C. L.: *Aptitude Testing*, 266.

[28] Henmon, *op. cit.*, 10.

4. The Standard or Probable Error of Estimate.

The third method of interpretation is improved by a modification which is less open to the objection that exceptional individual cases may extend the range of error abnormally. By the modified method the standard[29] or probable[30] error of estimate of one measure by another is determined. Condit[31] thus reports that the probable error of the estimated value of scholarship grades predicted by scores in the Thurstone Psychological Examination is ± .34 on a 5-point scale with a theoretical average of 3.00. However, the fact that the scale is 5-points does not give any indication of the real error involved. Most of the marks may have been 3's and the variation in the distribution small, in which case, a P. E. (est.) of .34 would be comparatively large. On the other hand, with wide variation in the distribution and with all marks from 1 to 5 liberally used, a similar P. E. (est.) of .34 would be comparatively small.

It is evident, therefore, that a more desirable method of indicating predictive value must take into account the variation in the distribution in which the forecasted score is located. This has been done by Crawford,[32] the Director of Personnel Study at Yale, in stating that 90 per cent of all predictions made by combining entrance examination averages, school records and scores in the scholastic aptitude test, fell within a range of ± 1.00 standard deviation on the marking scale.

5. Hull's Predictive Efficiency Index.

Probably the simplest and most significant measure of predictive value is the per cent reduction in the standard error of estimate due to the use of the predictive index. The reduction in error when no index is used is zero. In other words, when no index is used, pure chance placement is assumed. This measure, described by Hull,[33] is known as the predictive efficiency index,

---

[29] σ (est) = $\sigma_2 \sqrt{1 - r_{12}^2}$.

[30] P. E. (est.) = .6745 $\sigma_2 \sqrt{1 - r_{12}^2}$

[31] Condit, *op. cit.*, 333.

[32] Crawford, A. B.: "Forecasting Freshman Achievement." *School and Society*, 31: 125–132. January 25, 1930.

[33] Hull, *op. cit.*, 268.

E. E is determined by dividing the standard error of estimate by the standard deviation of the distribution in which the forecast is made, and subtracting the resulting fraction from 1.00. The answer is then expressed as a per cent. Since the standard error of estimate is derived from the coefficient of correlation it is usually more convenient to determine E directly from the equation:[34]

$$E = 1 - \sqrt{1 - r^2}$$

The devices for measuring predictive efficiency as they have been described above are:

1. Pearson product-moment correlation coefficient.
2. The per cent of the total number of cases correctly sectioned.
3. The ranges of achievement at successive placement levels.
4. The standard or probable error of estimate.
5. Hull's predictive efficiency index, E.

It was necessary to select those which would be used in this study. The selection was based upon four criteria. The measure of predictive efficiency used in this study should:

1. Be independent of a number of variables whose values may or may not be known. Among these unknown variables may be the number of sections into which the entire group is divided, the distorting effect of exceptional individual cases, the unknown scale values of the units in which errors are reported, and differences in the variations of scores within groups.
2. Be capable of combination with other indices in the multiple regression equation to increase the validity of prediction.
3. Be intelligible to school administrators without a detailed knowledge of statistics.
4. Be comparable with measures used in other studies already reported.

The first and third conditions seem to be satisfied best by Hull's predictive efficiency index, E, which is not dependent upon the number of groups sectioned, nor seriously affected by errors introduced in considering only exceptional cases, nor open to the objection of ambiguity because of differences in the

[34] *Ibid.*

variation of the criterion measures. E, as a measure of efficiency, expressed as a per cent of perfect prediction, is comparable with measures of efficiency in fields other than the prediction of achievement and, consequently, is to that degree more intelligible. The second and fourth conditions are best satisfied probably by the Pearson product-moment coefficient of correlation, r. To make the results of this study as usable as possible, results will be given in terms of both correlation coefficients and Hull's indices of predictive efficiency.

III. *The Relative Value of IQ's in Predicting Specific Subject Achievement in Different Parts of the IQ Range*

Careful examination of the literature revealed only one research study in this field. This was reported by Branson.[35] In the Long Beach High School, Long Beach, California, the freshmen were grouped in English according to their scores in the Otis test. Branson reports that this scale appeared to select inferior pupils better than superior ones, although it was fairly efficient in both respects. The dearth of evidence in this special field indicates the need for research.

IV. *The Extent of Variation in Individual Achievement in Different Subjects*

Profile charts reporting the relative achievement scores of individuals in different subjects show some individuals with all scores at relatively the same level. For other individuals, the scores in different subjects vary widely from each other. The average amount of this variation in the scores of individuals is measured in this study.

The published research articles dealing with this phase of the problem, and located by the author, are limited to six in number, three of which are pertinent. Studies by Kelley, Gates, and Brown respectively reveal the facts (1) that individual irregularities in achievement are significant, (2) that their discovery has cast doubt upon certain educational practices and (3) that

[35] Branson, E. P.: "An Experiment in Arranging High School Sections on the Basis of General Ability." *Journal of Educational Research*, 3: 53-55. January, 1921.

they tend to be greater for individuals with the higher mental ability test scores. This last statement is further substantiated by a research in psychology: Wallin[36] has found greater scatter in the specific performance of pupils in various tests of the Binet scale for normal than in those for sub-normal children. Pyle[37] finds significant unevenness in specific types of learning among freshmen at Detroit Teachers' College. He states:

"For four semesters I have taken a few of the very highest and a few of the very lowest ability students as determined by the intelligence test (Detroit Advanced Intelligence Test) and made extensive laboratory experiments upon them. In every case, I have found some of the low ability students excelling some of the high ability in certain types of learning. It is rare to find the same student either uniformly high or uniformly low in all learning experiments."[38]

Kelley[39] devised new statistical techniques to determine the reliability of the irregularities shown in individual scores for the separate tests of the Stanford Achievement Test. He concludes that a large part of the irregularities shown in the test profiles of 96 Palo Alto eighth grade pupils were statistically significant.

De Voss[40] using Kelley's techniques reached a similar conclusion as the result of a study of similar scores for 100 gifted children. In addition, the De Voss study applied other techniques and reached conclusions substantially in agreement with Kelley's. One finding was the determination of but slightly greater specialization of ability among the gifted children than among others. The difference was not great enough to be reliable.

Gates, after finding low intercorrelations among elementary school subjects, says,

---

[36] Wallin, J. E. W.: "Intelligence Irregularity as Measured by Scattering on the Binet Scale." *Journal of Educational Psychology*, 13: 140–151. March, 1922.

[37] Pyle, W. H.: "The IQ and the Individual." *School and Society*, 26: 788–790. December 17, 1927.

[38] Pyle, *op. cit.*, 790.

[39] Kelley, T. L.: "A New Method for Determing the Significance of Differences in Intelligence and Achievement Scores." *Journal of Educational Psychology*, 14: 321–333. September, 1923.

[40] De Voss, J. C.: "Specialization of the Abilities of Gifted Children." Ch. XII, 307–362, in Terman, L. M. et al., *Genetic Studies of Genius*, Vol. 1.

"The two most likely explanations of the low intercorrelations among school subjects are: (1) a specialization among the functions due primarily to original (inherited) aptitudes and (2) differences in the degree of possible achievement in each, depending upon the relative emphasis in teaching."[41]

He further criticizes the use of the accomplishment quotient because that measure is based on the assumption of little or no differences in the increase of special abilities.

Brown's[42] research study of 456 ten-, eleven-, twelve-, and thirteen-year-old Easton children offers evidence that relative variation in performances in the Haggerty Intelligence Test Delta 2, the Pintner Non-Language Mental Test, the Stenquist Mechanical Aptitude, and the Stenquist Assembly seemed to be greater for the bright than for the dull children.

Summarizing, it may be seen that no data have been presented regarding variation in subject achievements of students in the upper twelfth grade. However, data obtained from the records of elementary school children, 8th grade pupils, gifted children, and freshmen in a teachers' college show that significant individual variation exists among them in certain types of school achievement. It will be the problem of this study to present with some degree of definiteness the extent of individual variation at the upper twelfth grade level.

V. *The Ratio of the Average Standard Deviation of the Individual Scores to the Standard Deviation of the Most Complete Group Distribution*

In an effort to make the variation of scores around the individual means, as described in IV above, more nearly comparable with results for other studies, this variation is expressed as a fractional part of the variation of the total group in one subject, English.

Only one research is reported which attempts to measure the extent of individual variation as a ratio of the group variation. Hull[43] reports a study of the scores of individuals in a number

---

[41] Gates, *op. cit.*, 277.
[42] Brown, A. W.: *The Unevenness of the Abilities of Dull and Bright Children.*
[43] Hull, *op. cit.*, 47.

of trait tests. The variation of the trait scores of each individual measured from the average score of that individual was determined. These individual trait variations were compared with the variations of all individuals about the group mean score in each trait. From this study Hull concludes that

"the extent of trait differences within the average individual was found to approach rather closely the amount of difference found in a normal group in respect to any single trait."[44]

He indicates that, after allowing for errors in the data, individual trait variation is 80 per cent of group variation.

VI. *The Extent to Which Achievement in Specific Subjects Can Be Predicted from Average Achievement*

While a relatively large number of studies have attempted to determine the value of actual measures or of indices of general achievement in predicting specific achievement, none have been found which attempted to determine the probable limitations of this device. It is evident that the accuracy with which a test of mental ability can predict achievement in each of a number of subjects is limited by at least two conditioning factors:

1. The accuracy of prediction of average achievement by the mental ability test.
2. The correlation between average achievement and specific achievement.

This study furnishes evidences pertaining to the second factor.

---

[44] *loc. cit.*

CHAPTER III

Description of the Data

I. *The Testing Program*

The data used to determine the relationships between the IQ and specific subject achievements were obtained as part of the Cooperative Study of the Relations of Secondary and Higher Education. This study was sponsored by the Carnegie Foundation for the Advancement of Teaching and the Department of Public Instruction of Pennsylvania cooperating with the Board of Public Education in Philadelphia.

As a significant part of the Cooperative Study, Section II entitled "A Five-Year Cumulative Study Beginning with Twelfth-Grade Pupils in Secondary Schools" was initiated May 7 and 8, 1928. The Otis Self-Administering Test of Mental Ability Higher Examination, Form A, and a battery of eighteen achievement tests were administered to the graduating classes. The names of the tests and the numbers of pupils tested are listed below:

NAMES OF TESTS	NUMBERS OF PUPILS TESTED
Otis Self-Administering Test of Mental Ability (Higher Examination)	2449
Columbia Research English	2396
Columbia Research French	643
Columbia Research German	180
White Latin	527
Columbia Research Spanish	467
Ruch-Crossman Biology	281
Columbia Research Chemistry	727
Ruch-Popenoe General Science	649
Columbia Research Physics	433
Columbia Research Algebra	1940
Columbia Research Plane Geometry	1008
American Council Solid Geometry	89
American Council Trigonometry	160
Columbia Research American History	2370
American Council Ancient History	47
American Council Civics	250
American Council Economics	129
American Council General European History	164
Total Number of Pupil Tests	14909

The administration of the tests in each of the twelve senior high schools of Philadelphia was directed by an examiner appointed from the faculty of the school and supervised by the Division of Educational Research of the Philadelphia Public Schools. In order to secure, as far as possible, uniform conditions of testing, detailed directions[1] were furnished each examiner for the carrying out of the general program in addition to the printed manual for each test. Most tests were scored by a corps of 14 normal school graduates under supervision in the office of the Division of Educational Research of the Philadelphia Board of Education. The scoring of all sets of test papers was checked by having a minimum sampling of 10 per cent rescored by a person other than the original scorer. Where errors in scoring were frequent or constant the entire set of test papers was rescored. On account of the significance of the Otis test scores, each paper was scored twice by different persons.

An additional source of data was a "Pupil's Information Sheet" returned by the schools for each student tested. A copy of this blank is reproduced in Appendix B, page 141. Entries made by the 12th grade students on the "Pupil's Information Sheet" for chronological age and the duration and recency of instruction were used in this study. The average school mark, entered by the school authorities, was also used. The average school mark does not measure the status of students at the end of their high school course. However, the greater reliability of an average of several terms was used to make the marks more reliable than the marks of the last half year. It must be recognized, however, that the use of the average school mark is based upon the assumption that pupils vary less from term to term in their achievement than the subjective estimates of success, provided by the school marks at the end of the period of instruction, vary from "true" estimates.

In entering these averages, four marks, A, B, C, and D were used. This procedure was a departure from the obtaining practice in assigning marks in Philadelphia high schools, where there

---

[1] See Appendix A, pp. 135–139, for the general directions which are described under the caption "Procedure to be followed in the Examination of Twelfth Grade Pupils."

are five marks assigned, E, G, F, P, and V. P., the last two of which are "failing" ratings. In transmuting marks from the 5-point Philadelphia scale to the 4-point scale used in the statewide study, the following interpretation was given for each mark:

MARK FOR COOPERATIVE PROGRAM		MARK IN PHILADELPHIA HIGH SCHOOLS
A	=	either E or G
B	=	high F
C	=	low F
D	=	either P or V. P.

This plan of transmutation of school marks in effect tended to prevent restriction of range incident to averaging marks although the resulting distributions usually did not approximate the normal frequency distribution. Accordingly, a special scaling procedure[2] was required in order to utilize the average school marks reported on the "Pupil's Information Sheet" as a source of data for this study.

Summaries of the data in the form of master charts were made for each student. Distributions of test scores and average school marks were taken directly from these charts. The numbers of terms of instruction, and the numbers of terms which had elapsed since the last period of instruction were made the bases of the selection of the pupils whose records were studied. The chronological ages given on the "Pupil's Information Sheet" were used in the determination of IQ's.

II. *Validity and Reliability of the Subject Matter Tests*

It is believed that the official and cooperative nature of the testing program providing, as it did, for the testing of 89 per cent of the total enrollment[3] in the upper twelfth grades of the senior high schools of Philadelphia, added materially to the reliability of data. It is obvious, however, that in the administration of 14,909 pupil tests there would be some scores which

[2] Described in Ch. IV., pp. 52–54.

[3] Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 97, *Mental Ability of Philadelphia 12B Pupils as Measured by Scores in Otis Test of Mental Ability, Higher Examination*, 1928.

would not represent accurately the achievement levels of the students tested. Where such facts were known the test scores were not included in this study.

In order to provide suitable checks upon the validity of the masses of data analyzed, the assumption was made that test scores and school marks denote achievements with similar elements to the degree that the scores and marks correlate positively. On this basis no measures of specific achievement in biology were included in this study. Table I reports correlation coefficients between test scores and the average marks in the respective subjects. A correlation coefficient of .064 ± .053 is reported in Table I for scores in the Ruch-Crossman Biology Test and average school marks in the subject. Positive correlation is desirable but high coefficients should not be expected if the test scores are to add any validity to the average school mark.[4] It is only to the extent that test scores and school marks measure different kinds of achievement status that any additional value is to be attached to school marks supplemented by standardized test scores. Perfect correlation between test scores and school marks would indicate that either could replace the other. On the other hand, extremely low or zero correlations between these two criteria of school success raise questions as to the reliability of the marks or the scores. Rather than include such questionable data, no measures were used for the subject indicated above.

Table I shows that all correlation coefficients are positive. With the exception of the correlation coefficient between the Ruch-Crossman Biology Test and average school mark in biology, all coefficients exceed .200 and are lower than .600, with the median at .353. In the case of five subjects in which

---

[4] A study of the factors which influence grades in the Philadelphia junior and senior high schools shows that many factors, other than the pupil's achievement in the subject, were influential. Among these other factors were: attitude toward the subject (teacher?), interest, cooperation in class projects, whether or not the pupil worked up to the limit of his ability, etc. See Philadelphia Board of Education, Division of Educational Research and Results, Bulletins 194, 195—*Summary of Subject Promotion Rates, Junior and Senior High Schools*, January, 1931. It is evident that if the tests are valid measures of achievement in specific subject fields, high correlation coefficients between school marks and test scores are not be be expected.

the numbers of students were small, the coefficients of correlation do not exceed three times their probable errors. These subjects are: German, biology, solid geometry, ancient history and economics.

TABLE I

CORRELATION COEFFICIENTS BETWEEN SCORES IN THE SUBJECT-MATTER TESTS AND AVERAGE SCHOOL MARKS IN THE SUBJECTS

NAMES OF TESTS	GROUP TESTED		NUMBER OF STUDENTS	r	P. E.$_r$
	NUMBER OF TERMS INSTRUCTION	NUMBER OF TERMS ELAPSED SINCE THE LAST SCHOOL MARK IN THE SUBJECT			
Columbia Research English	8	0	2227	.452	.013
French	6	0	169	.372	.048
German	6	0	16	.235	.150
White Latin	8	0	313	.587	.025
Columbia Research Spanish	4	4	142	.353	.050
Ruch-Crossman Biology	2	4	162	.064	.053
Columbia Research Chemistry	2	0	258	.348	.037
Physics	2	0	182	.259	.052
Algebra	4	1	628	.466	.022
Plane Geometry	2	3 and 4	604	.553	.019
American Council Solid Geometry	1	3	51	.233	.090
Trigonometry	1	0	125	.450	.049
Columbia Research American History	2	2	1759	.343	.016
American Council Ancient History	2	6	31	.282	.114
Civics	2	0	133	.353	.052
Economics	1	1	45	.209	.096
General European History	2	4	98	.323	.060
Median coefficient of correlation excluding that for Biology				.353	

For the interpretation of the correlation coefficients reported in Table I, it is essential to consider certain factors in addition to lack of test validity which tend to reduce the possible magni-

tude of these coefficients. The correlation coefficients between the test scores and the school marks are limited by (a) the reliabilities of the measures, (b) the differences in variation of achievement in the test groups, and (c) the coarseness of the interval grouping in the distributions of the school marks.[5]

It is probable that each coefficient reported in Table I would be higher and therefore indicate closer relationship between the two measures of achievement if the measures were perfectly reliable, if the variations of the scores for the measured groups in each subject were uniformly those of the total group in all subjects, and if school marks could be distributed into ten or twelve categories instead of only the four indicated in the directions for reporting data on the Pupil's Information sheet.[6] The limiting effects described above would probably prevent correlation coefficients reported in Table I from rising above an estimated maximum[7] of .81 except through chance, even though the achievements measured by the test scores and the school marks were identical.

However, similar inaccuracies of measurement exist in most determinations of the validity coefficients of tests and correlation coefficients are reduced accordingly. Comparisons are usually made with similarly derived correlation coefficients

---

[5] Formulae and descriptions of the application of these "corrections" ar found as follows:
Reliability (attenuation)—Hull, C. L.: *op. cit.*, 237–243. Variation of score distributions of groups—see pp. 69–70 of this study. Coarseness of interval grouping (Sheppard's correction)—Kelley, T. L.: *Statistical Method*, 169.

[6] See Appendix B, p. 141.

[7] The application of the correction for attenuation alone is illustrated below:
Assume average reliabilities (r_{11} and r_{22}) of .90 and .74 for achievement test scores and school marks respectively. (See Table IV, p. 47 and pp. 46–49.) If the achievements measured by test scores and school marks were identical, the "true" correlation coefficient, $r_{\infty\infty}$, would equal 1.00. Solving the Spearman attenuation formula for the "actual" correlation coefficient, r_{12}:

$$r_{\infty\infty} = \frac{r_{12}}{\sqrt{r_{11}} \cdot \sqrt{r_{22}}}$$

$$r_{12} = r_{\infty\infty} \cdot \sqrt{r_{11}} \cdot \sqrt{r_{22}}$$

$$r_{12} = \sqrt{.90} \cdot \sqrt{.74} = .81$$

The coarseness of interval grouping would further reduce this coefficient.

TABLE II

VALIDITY COEFFICIENTS FOR SUBJECT-MATTER TESTS USED IN THE COOPERATIVE PROGRAM

NAME OF TEST	TEST SCORES CORRELATED WITH	GROUP TESTED	REPORTED IN	r
Columbia Research				
English	College Grades	Columbia Freshmen	Author's Manual	Aver. .50
French	College Grades		Author's Manual	.70
	Regents Exam. Marks	2nd Year Students	Author's Manual	.71
	Regents Exam. Marks	3rd Year Students	Author's Manual	.72
	Regents Exam. Marks	4th Year Students	Author's Manual	.58
German	College Grades		Author's Manual	.70
	Regents Exam. Marks	2nd Year Students	Author's Manual	.72
	Regents Exam. Marks	3rd Year Students	Author's Manual	.76
	Regents Exam. Marks	4th Year Students	Author's Manual	.72
Spanish	College Grades		Author's Manual	.70
	Regents Exam. Marks	2nd Year Students	Author's Manual	.55
	Regents Exam. Marks	3rd Year Students	Author's Manual	.65
	Regents Exam. Marks	4th Year Students	Author's Manual	.71
Chemistry	College Marks		Author's Manual	.56
Physics	College Marks		Author's Manual	about .70
	Regents Exam. Marks		Author's Manual	about .60
Plane Geometry	Secondary School Marks	800 Cases in Eastern Secondary Schools	Author's Manual	Aver.55 .50
	"Success in college math."			
American Council				
Solid Geometry	School Marks	51 Independent School Grads.	Ed'l. Records Bur.	Aver. .552
Trigonometry	School Marks	20 Independent School Grads.	Ed'l. Records Bur.	.685
Columbia Research				
American History	Final Grades in High School	622 DeWitt Clinton High School Students	Author's Manual	.822
	Regents Exam. Marks		Author's Manual	.469
American Council				
General European History	School Marks	87 Independent School Grads.	Ed'l. Records Bur.	.765

between test scores and school marks. "Corrections" of the coefficients reported in Table I for differences in the reliabilities of the measures, differences in the variations of scores in the distributions, and the coarseness of the interval groupings are not desirable in this instance.

In order to afford comparisons for some of the correlation coefficients reported in Table I, determined for groups of the upper twelfth grade of Philadelphia high schools in May 1928, the authors' validity coefficients for the same tests are given in Table II. Most coefficients are larger than those in Table I. The fact that the test scores do not correlate so highly with the school marks for the groups in this study as they do for the groups reported in Table II serves to emphasize the need for the inclusion of the test scores in any comprehensive criterion of subject success. In other words, both school marks and test scores measure achievements which are probably different to a considerable degree.

Additional evidence of the validity of scores in four tests as measures of achievements in the subject is furnished by Table III. Table III reports for the data of the Cooperative Program correlation coefficients between the scores in four tests and the number of terms of instruction in the respective subjects. In the four subjects—French, German, Latin, and algebra,—the program of studies was flexible enough to permit different students to be instructed in the subjects for varying numbers of terms. In some cases, students of low achievement levels spent more time in the subject through failure. The effect of this was probably to reduce the obtained correlation coefficients. However, the correlation coefficients between test scores and the number of terms instruction in Table III exceed in every case coefficients for the same tests reported in Table I for test scores and average school marks.

This would seem to indicate that the validity of the tests is higher in measuring those kinds of achievement which distinguish the students with longer periods of instruction from those who have apparently spent less time in study of the subject. In other words, they seem to measure that kind of specific subject ability which results from spending assigned school time in the given fields.

TABLE III

CORRELATION COEFFICIENTS BETWEEN SCORES IN SUBJECT-MATTER TESTS AND THE NUMBER OF TERMS INSTRUCTION

NAME OF TEST	NUMBER TESTED	r	P. E.$_r$
Columbia Research French	715	.752	.011
Columbia Research German	169	.495	.040
White Latin	446	.672	.018
Columbia Research Algebra	1730	.583	.012

The schedule of examinations (Appendix A, pp. 135–139) shows that all standardized subject-matter tests used required at least one hour for administration. Most tests required one and one-half hours, while one, the Columbia Research English test, required two hours. Most tests consisted of more than 150 question items. Accordingly, high reliability coefficients might be expected. Table IV, reporting reliability coefficients for the tests, shows that such is the case. All reliability coefficients reported are above .90 except those for biology, chemistry, physics, and algebra, which are above .80.

III. *Reliability of the Average School Marks*

Reliability coefficients for the teachers' marks, where these coefficients can be determined, do not ordinarily approach the magnitude of the coefficients reported in Table IV for the standardized tests. This is but another way of saying teachers' marks are not so objective as test scores. Evidence for this is furnished by numerous studies[8] of the reliability of school marks. It should be recognized, however, that the startling revelations of lack of reliability of school marks discovered in the earlier

[8] A bibliography of 79 titles under the caption "School Marks" is provided by Odell, C. W.: "A Selected Annotated Bibliography Dealing with Examinations and School Marks." *University of Illinois Bulletin*, 24: 32–41. January 15, 1929.

TABLE IV

Reliability Coefficients Reported for Tests Used in the Cooperative Program

NAME OF TEST	METHOD OF DERIVATION	GROUP TESTED	REPORTED IN	r
Otis Mental Ability	Form A with Form B	Grades 7 to 12, 125 & 128 cases	Author's Manual	.921*
Columbia Research				
English	Spearman Brown[9]	100 entering college students	Author's Manual	.965
French	Spearman Brown[9]		Author's Manual	.96
German	Spearman Brown[9]		Author's Manual	.96
Spanish	Spearman Brown[9]		Author's Manual	.97
Ruch-Crossman				
Biology			Kelley, T. L.: *Interpretation of Educational Measurements*, 332	.82
Columbia Research				
Chemistry	Spearman Brown[9]		Author's Manual	.87
Physics	Spearman Brown[9]	300 entering college	Author's Manual	.90
	Spearman Brown[9]	575 entering college	Author's Manual	.86
Algebra	Form A with Form B		Author's Manual	.86
Plane Geometry	Spearman Brown[9]		Author's Manual	.93
American History	Spearman Brown[9]		Author's Manual	.91
American Council				
Economics			Odell, C. W.: *Educational Measurements in High School*, 297	.83

* See also p. 119.

[9] The Spearman-Brown formula is applied to chance halves of a single form of a test. The technique is given in Holzinger, K. G.: *Statistical Method in Education*, 169; and in Otis, A. S.: *Statistical Method in Educational Measurements*, 263, 264.

studies[10] have not been borne out by the findings of the more recent studies.[11] This may be due partly to the employment in the earlier studies of techniques which may be classed as highly artificial and to the establishment of conditions considerably removed from actual school marking practice. Higher reliability may also have resulted from efforts to improve marking systems.

An estimate of typical reliability coefficients for school marks may be had by the correlation of first and second term marks. Toops,[12] Wood,[13] Symonds,[14] Crawford,[15] and the College Entrance Examination Board[16] report such correlations coefficients: .66, .69, .73, .80, .83 respectively.

While these coefficients are not "true" reliability coefficients, *i. e.*, obtained from the correlation between different measures of the same thing, they establish values for the reliability of school marks which are probably lower than self-correlations would be, because they are reduced by such variables as the fluctuations in pupil effort.

In an effort to approximate more nearly a self-correlation of school marks, the Division of Educational Research of the Philadelphia Public Schools has correlated the average marks for odd-numbered terms of instruction in French, German, and Spanish with the average mark for the even-numbered terms. For

---

[10] Kelly, F. J.: *Teachers' Marks, Their Variability and Standardization.*
Rugg, H. O.: "Teachers' Marks and the Reconstruction of the Marking System." *Elementary School Journal*, 18: 701–719. May, 1918.
Starch, D.: "Marks as Measures of School Work," and "A Sample Survey of the Marking System in a High School." Chs. II and III of *Educational Measurements.*
Starch, D. and Elliott, E. C.: "Reliability of Grading Work in English, Mathematics, History." *School Review*, 20: 442–457. 21: 254–259; 676–681. September, 1912, April and December, 1913.

[11] Bolton, F. E.: "Do Teachers' Marks Vary as Much as Supposed?" *Education*, 48: 23–39. September, 1927.
Symonds, P. M.: *Measurement in Secondary Education*, 426.
Crawford, A. B.: "Forecasting Freshman Achievement." *School and Society*, 30: 125–132. January 25, 1930.

[12] Toops, H. A.: "The Use of Intelligence Tests by Universities." *School and Society*, 24: 603. July 17, 1926.

[13] Symonds, P. M.: *Measurement in Secondary Education*, 426.

[14] *Ibid.*

[15] Crawford, A. B.: "Forecasting Freshman Achievement." *School and* Society, 30: 125–132. January 25, 1930.

[16] *College Entrance Examination Board Annual Report, 1928.* 199.

example, the average marks in French (1) and French (3) were correlated with the average marks in French (2) and French (4). This technique reduces the variation in school marks which can be attributed to factors other than the lack of objectivity in the teachers' ratings. Of 70 such correlation coefficients the median[17] is .74. Accordingly, the average reliability of an average mark for two terms of instruction in these modern languages is probably no less than .74. For some subjects included in this study the marks for more than two terms of instruction were averaged,[18] the reliabilities should be correspondingly higher.

IV. *Validity and Reliability of the Measures of Mental Ability*

The tests in this program were administered by teachers under the direction of an examiner in each school. Under these circumstances it was desirable that the tests be simple in administration, without short separately timed sub-tests. This is particularly true of the test of mental ability because unreliability in the measure of mental ability would affect all relationships of mental ability scores with those of other tests. Accordingly, it is believed that the selection of a self-administering test such as the Otis Test is justified by the greater uniformity of administration. In Table IV a reliability coefficient[19] of .921 for this 30-minute examination of mental ability is reported.

Scores in the Otis Self-Administering Test of Mental Ability, Higher Examination, were studied as predictive indices of average achievement. Their value for this purpose may be estimated from the validity coefficients in Table V. In Table V, correlation coefficients between scores in the Otis test and scores in other tests of mental ability are higher than the correlation coefficients for Otis test scores and criteria of school success. It is probable that the lower correlation coefficients are due not only to differences in the functions measured by tests of mental ability

---

[17] Philadelphia Board of Education, Division of Educational Research. Bulletin 201, *American Council Alpha Test Scores and Final Marks in French (4), German (4) and Spanish (4)*. January 1931.

[18] See Table I, p. 42.

[19] A wide grade range, i. e., over grades 7 to 12, causes this coefficient to be an exaggerated estimate for grade 12 alone. See p. 119.

TABLE V

VALIDITY COEFFICIENTS OF THE OTIS SELF-ADMINISTERING TEST OF MENTAL ABILITY, HIGHER EXAMINATION

CRITERIA:	REPORTED BY	GROUP TESTED	r
1. OTHER MENTAL ABILITY TESTS			
Otis Self-Administering Intermediate	*Otis Test Manual*, (1922)	100 cases in grades 7–9	.842 (av.)
Terman Test	Kornhauser, A. W.: *Journal Educational Research*, 16:346	116 cases first year undergraduate school of business	.82
Miller Intelligence	Bureau Public Personnel Administration. *Public Personnel Studies*, 7: 75		.73
Haggerty Intelligence Delta 2	Bureau Public Personnel Administration. *Public Personnel Studies*, 7: 75		.73
Stanford-Binet	Bureau Public Personnel Administration. *Public Personnel Studies*, 7: 75		.66
Composite consisting of Miller, Haggerty Delta 2, Terman, Army Alpha, Trabue's Mentimeters, and Otis	Miller, W. S.: *National Society for the Study of Education 21st Year Book:* 216	High school freshmen	.841
2. SCHOOL SUCCESS			
Average of five subject criterion scores	this study p. 17	183 members of senior high school graduating class	.617
First year marks in business school	Kornhauser, A. W.: *Journal Educational Research*, 16:345	116 cases in first year undergraduate school of business	.49
"Scholarship"	*Otis Test Manual*, (1922)	Grade 11—240 cases	.55
"Scholarship"	*Otis Test Manual*, (1922)	Grade 12—204 cases	.57
"Scholarship"	*Otis Test Manual*, (1922)	157 high school freshmen, Oakland, California	.59

and by criteria of school success, but also to somewhat lower levels of reliability for these criteria[20] in comparison with test scores. Thus W. S. Miller, of the University of Minnesota, states,

"Higher correlations between mental tests and school marks than are now obtained can not be expected until marks are based more exclusively on achievement."[21]

The validity coefficients reported in Table V for the Otis test are higher than similar coefficients in Tables I and II for the subject-matter tests. It is likely, therefore, that Otis test scores are at least as valid, and probably more valid indices of average achievement in all subjects as most of the subject-matter test scores are of the specific achievement. An Otis test validity coefficient of .617 for the correlation with an average of five criterion scores was determined for this investigation. Comparison of this coefficient of .617 with other correlation coefficients in Table V shows that the validity of the Otis test for this investigation is at least as great as that for other investigations. Since the coefficient .617 for this study exceeds each of the other four coefficients in Table V, there is the possibility that the validity of the Otis test may be correspondingly greater.

As these data have been described, it will be noted that they include for a total of 2449 students or 89 per cent of the enrollment of the graduating class of the Philadelphia senior high schools for June 1928:

1. Average school marks.
2. Standardized achievement test scores.
3. Scores in the Otis Self-Administering Test of Mental Ability, Higher Examination.

---

[20] See Table IV, p. 47, and pp. 46-49, 62 for the reliabilities of test scores, school marks, and for the Spearman correction for attenuation.

[21] Miller, W. S., *op. cit.*, 211.

CHAPTER IV

The Measures Used

I. *The Measure of Specific Subject Achievement: The Subject Criterion*

The measures of specific subject achievement should be valid, accurately scaled, and comparable. The data furnished two measures of specific subject achievement: (1) average of the school marks in the individual subjects, and (2) subject-matter test scores. These are measures of achievement similar to the degree of positive correlation (Table I). However, they were not scaled accurately, nor were they expressed in comparable units. Accordingly, the problem of providing adequate criterion measures of subject achievement involves combining the average school mark and the subject-matter test score of each individual into a single scaled criterion score for each of the subjects considered.

The school mark and the test score distributions were first changed to centile distributions. The centile values of marks and scores were then translated into sigma measures in units of one-tenth of the standard deviation of the distribution. These equal distances along the base line of the normal distribution have been termed "linear scores" following the practice of Hull.[1] The linear score for the test score was added to a similar unit for the school mark to form the subject criterion score.

The method may be illustrated by data for the criterion scores in English. Distributions of school marks and test scores were obtained from the scatter diagrams from which the correlation coefficients between test scores and school marks were computed. In the distribution of school marks, that centile scale value of each mark was taken which represented the median of the interval for the mark. Table VI reports the distribution of school marks in English. Thus, for the lowest mark in the scale, D, a total frequency (f) of 27 is recorded. Accordingly,

[1] Hull, C. L., *op. cit.*, 386–390.

for the centile value of D, one-half the frequency, or 13.5, is divided by the total number of cases, 2227. The resulting per cent is 0.6. The frequency (f) reported in Table VI for the mark C is 839. Accordingly, the number of cases in the distribution with marks exceeded by a mark of C at the median of the C interval is 27, the frequency for D, plus 419.5, or one-half the frequency for C. This total, 446.5, divided by 2227, the total number of cases in the distribution, and multiplied by 100 is 20.0, the centile value of the mark C. In a similar manner centile values were determined for marks of B and A.

Table VII gives base line, or sigma values, for the per cent areas of the normal frequency surface. These base line or linear values are expressed in tenths of the standard deviation distance from a mean arbitrarily chosen as 50. Thus, a point one standard deviation above the mean is 50 plus 10 or 60 linear score. Similarly, a point one standard deviation below the mean is 40 linear score, two standard deviations below is 30 linear score, etc. Thus, in Table VII, the linear score for the centile 0.6 is 25, centile 20.0 is 42, etc. These linear score values are given for each school mark in the right-hand column of Table VI.

TABLE VI

TRANSMUTATION OF SCHOOL MARKS IN ENGLISH TO LINEAR SCORES

SCHOOL MARK	f	NUMBER EXCEEDED PLUS ONE-HALF THE NUMBER ATTAINING	CENTILE	LINEAR SCORE
A	566	1944	87.2	61
B	795	1263.5	56.8	52
C	839	446.5	20.0	42
D	27	13.5	0.6	25
Total N =2227				

Figure 1 illustrates the transmutation of school marks in English to linear score units. The four areas in the normal frequency surface represent the per cents of students receiving each school mark, D, C, B, and A. These areas are bisected by vertical lines representing the centiles of the distribution reported

in Table VI as medians in the reported frequencies for each of the four intervals. The points of intersection of these vertical lines with the base line determine the linear or sigma score values for the school marks.

TABLE VII

LINEAR SCORE VALUES OF CENTILES[2]

CENTILE	LINEAR SCORE	CENTILE	LINEAR SCORE	CENTILE	LINEAR SCORE
0.1	19	31	45	70	55
0.2	21	32	45	71	56
0.3	23	33	46	72	56
0.4	23	34	46	73	56
0.5	24	35	46	74	56
0.6	25	36	46	75	57
0.7	25	37	47	76	57
0.8	26	38	47	77	57
0.9	26	39	47	78	58
1	27	40	47	79	58
2	29	41	48	80	58
3	31	42	48	81	59
4	32	43	48	82	59
5	34	44	48	83	60
6	34	45	49	84	60
7	35	46	49	85	60
8	36	47	49	86	61
9	37	48	49	87	61
10	37	49	50	88	62
11	38	50	50	89	62
12	38	51	50	90	63
13	39	52	51	91	63
14	39	53	51	92	64
15	40	54	51	93	65
16	40	55	51	94	66
17	40	56	52	95	66
18	41	57	52	96	68
19	41	58	52	97	69
20	42	59	52	98	71
21	42	60	53	99	73
22	42	61	53	99.1	74
23	43	62	53	99.2	74
24	43	63	53	99.3	75
25	43	64	54	99.4	75
26	44	65	54	99.5	76
27	44	66	54	99.6	77
28	44	67	54	99.7	77
29	44	68	55	99.8	79
30	45	69	55	99.9	81

[2] These values are taken from the K-W table in Kelley, T. L.: *Statistical Method.* Appendix C, 373–385.

The method of transmuting raw scores in the Columbia Research English Test to comparable linear measures is quite similar to that just described for school marks in English. However, the greater number of step intervals made some modification of the procedure desirable. In this case the lowest score in each interval was taken as exceeding the cumulative total of all the scores in the intervals below it. These cumulative totals were converted to centiles and then to linear scores in the method above described for school marks. Table VIII illustrates

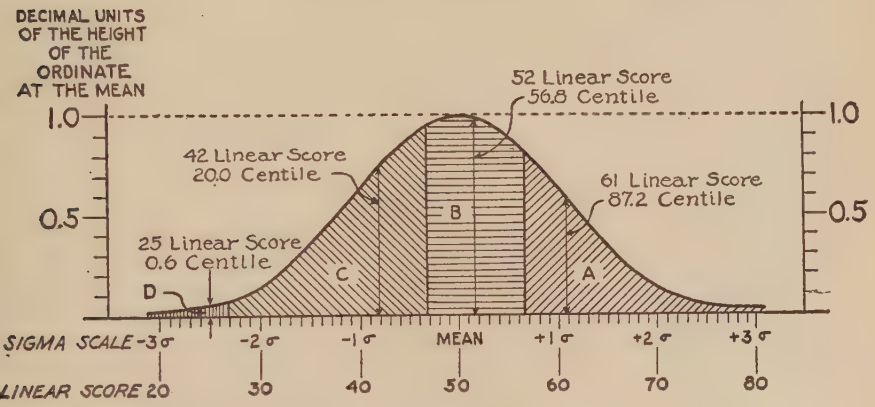


FIGURE I

Transmutation of School Marks in English To Linear Score Units, Showing the Frequency Surfaces of the Normal Frequency Distribution for Each School Mark and its Transmuted Linear Score

the method. In the column at the left, each interval is given. In the frequency (f) column, there is the number of cases in the interval. Thus, a frequency of 2 beside interval 50–69 in Table VIII indicates that 2 cases were below 70. Thus a score of 70.0 exceeded 2–2227ths or 0.1 per cent of the distribution. The score of 70.0 is located at the 0.1 centile and accordingly is assigned a linear score value of 19 from Table VII. Similarly, were determined a linear score value of 27 for a raw score of 90.0, a linear score of 36 for a raw score of 110.0, etc.

The right-hand column of Table VIII reports a linear score for each of the ten raw scores which are lowest in their respective intervals. These ten raw scores cover a range from 70 to 250. By plotting these ten values as in Figure 2, but on a larger scale, it was possible to determine the intermediate values by interpolation. A few scores at both ends of the range were determined by extension of the curve.

TABLE VIII

TRANSMUTATION OF TEN COLUMBIA RESEARCH ENGLISH TEST SCORES TO LINEAR SCORES

RAW SCORE	f	CUMULATIVE TOTAL FREQUENCY	LOWEST SCORE IN INTERVAL	
			CENTILE	LINEAR SCORE
250–269.....	4	2227	100.0	
230–249.....	10	2223	99.8	79
210–229.....	51	2213	99.3	75
190–209.....	150	2162	97.3	69
170–189.....	324	2012	90.5	63
150–169.....	535	1688	76.0	57
130–149.....	574	1153	52.0	51
110–129.....	399	579	26.0	44
90–109.....	150	180	8.1	36
70– 89.....	28	30	1.3	27
50– 69.....	2	2	0.1	19

In this way comparable linear scores were obtained for both school marks in English and Columbia Research English Test scores. The two linear scores were added to form the criterion score for each student in English.

This technique involves the following assumptions:

1. That the same group of students (but not necessarily the same individual students) exhibit the same mean level and variation of performance whether measured in the scale of the school marks or that of the test scores.

2. That, as a measure of subject achievement, a given centile position in the distribution of school marks was of equal value to a similar centile position in the distribution of test scores.

3. That distributions of true measures of both were of normal

frequency, no matter what degree of abnormality was shown in either the actual marks or the actual scores.[3]

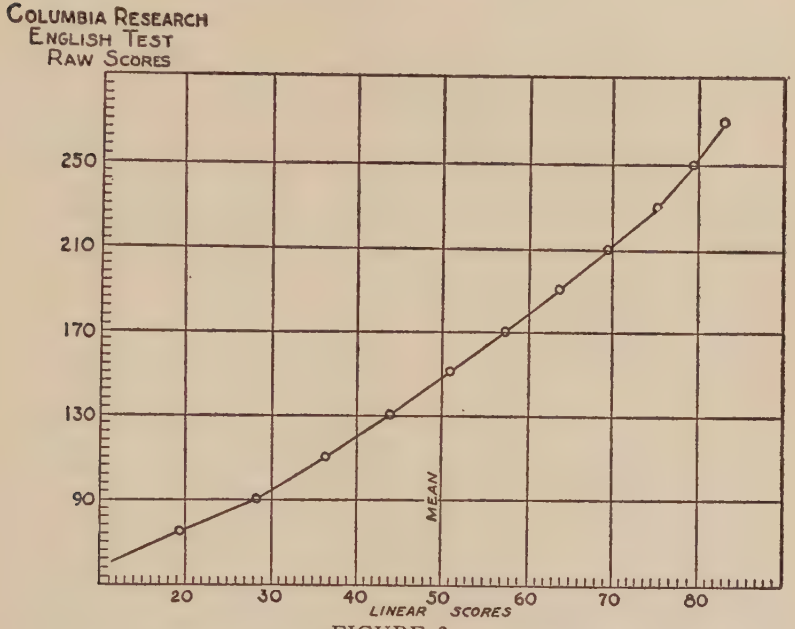


FIGURE 2

INTERPOLATION OF LINEAR SCORE VALUES FOR RAW SCORES IN COLUMBIA RESEARCH ENGLISH TEST

4. That the non-identical elements in test scores and school marks are of equal weights in determining total achievement in each subject. This fourth assumption cannot be proved. However, until it is known definitely what the non-identical elements are and what their relative importance is, it would seem that the most logical procedure would be based upon equal weighting rather than on some other arbitrary weighting without evidence. For example, it may be argued that since the average school mark was the result of four years of testing and observation, it

[3] These three assumptions underlie much of all measurement technique. See McCall, W. A.: *How to Measure in Education*, 249–306 for a discussion of the fundamental assumptions.

should be a more valid measure of achievement than a test score dependent upon performance during an hour and a half. However, consideration of the comparative unreliability of school marks should prevent weighting them heavily. On the other hand while the test scores may have high reliability coefficients, they leave some phases of achievement unmeasured.

The nature of these assumptions can be more carefully studied by the graphic method. Figures 3, 4, 5, 6, 7, and 8 showing, for English, distributions of school marks and test scores, distributions of linear values for the marks and scores, and the distribution of the composite criterion scores, enable the steps in the process to be more clearly identified. Figure 3 shows the distribution of average school marks in English. The base line of Figure 3 is scaled as if A were above B as much on the achievement scale as B is above C or C is above D. Figure 4 shows the same frequency distribution for the linear or sigma values of the average school marks. In Figure 4 the base line shows the relative distance between the average school marks is not uniform if the true distribution of achievement is assumed to be of normal frequency. Figure 5 gives the frequency of scores in each interval of the Columbia Research English Test. Figure 6 reports the same frequencies on a base line of linear scores. Figure 7 shows on a single base line the distribution of linear score values for average school marks, as in Figure 4, and the distribution of linear scores values of test scores as in Figure 6. Figure 8 shows the distribution of criterion scores resulting from the addition of the linear or sigma score values for each individual's school mark and test score. These figures show that the scaling and combining procedure has resulted in changing the abnormal original distribution of average school marks into one of approximately normal frequency. On the other hand, the greater dispersion of the test scores has been reduced in the averaging with school marks. In so far as the distribution of criterion scores represents more normal frequencies than that of the average school marks and less dispersion than the distribution of test scores they are probably more accurately scaled and more valid measures of achievement than either the raw scores or school marks alone.

It will be noted that the distributions of linear scores will have a mean of 50 in each case. Accordingly, the combined distribution (the English criterion score distribution) should have a mean of 50 plus 50, or 100, within the errors due to abnormal

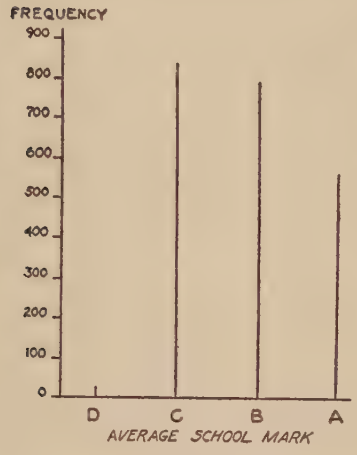


FIGURE 3
DISTRIBUTION OF AVERAGE SCHOOL MARKS IN ENGLISH

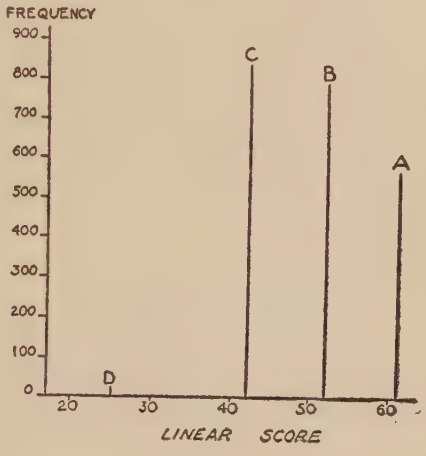


FIGURE 4
DISTRIBUTION OF LINEAR SCORE VALUES OF AVERAGE SCHOOL MARKS IN ENGLISH

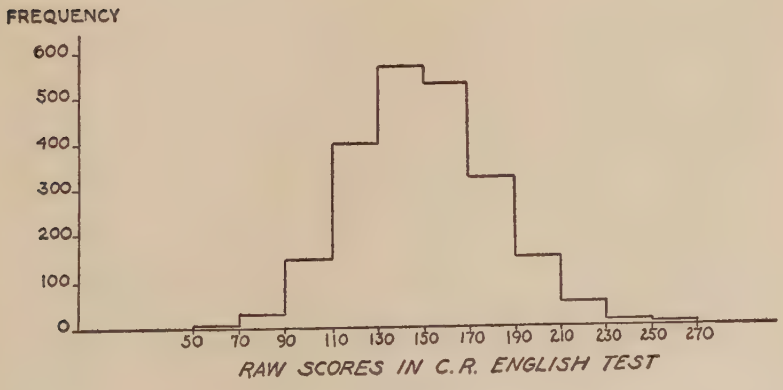


FIGURE 5
DISTRIBUTION OF RAW SCORES IN COLUMBIA RESEARCH ENGLISH TEST

distributions of school marks and test scores, and to the elimination of the decimals, and grouping into intervals. In this way, a mechanical check was provided for the arithmetical accuracy of determining the linear score values and the transmutation

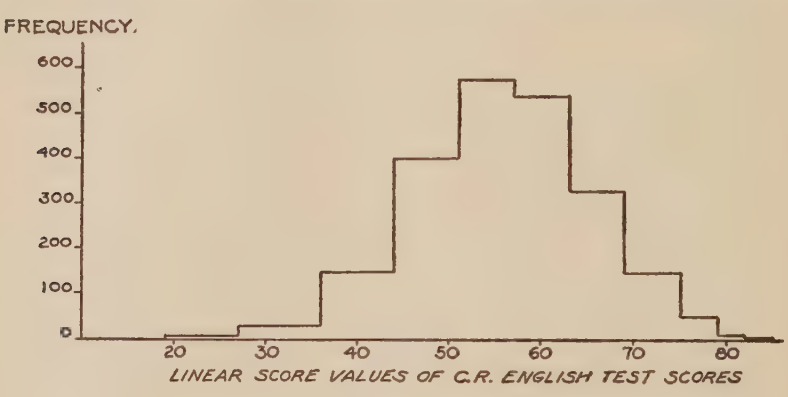


FIGURE 6
DISTRIBUTION OF LINEAR SCORE VALUES OF COLUMBIA RESEARCH ENGLISH TEST SCORES

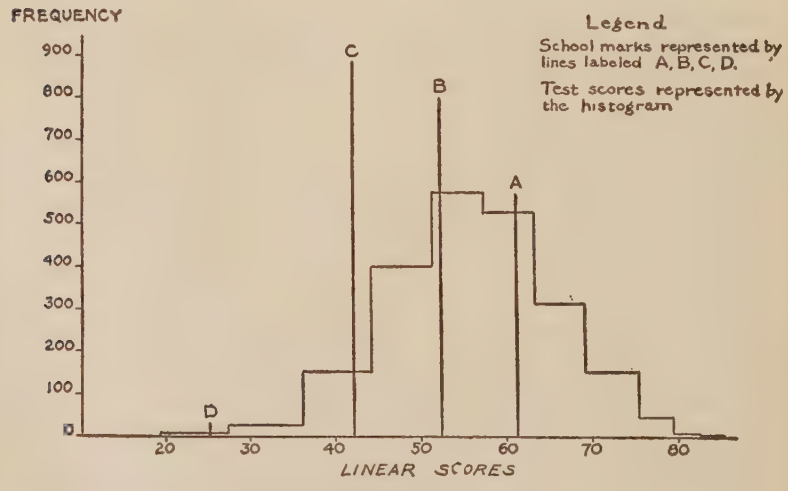


FIGURE 7
SUPERIMPOSED DISTRIBUTIONS OF LINEAR SCORE VALUES OF AVERAGE SCHOOL MARKS AND TEST SCORES IN ENGLISH

of the raw scores. The English criterion mean score, 99.4, indicated the probability that there had been no serious errors in the calculation of the transmuted scores for English.

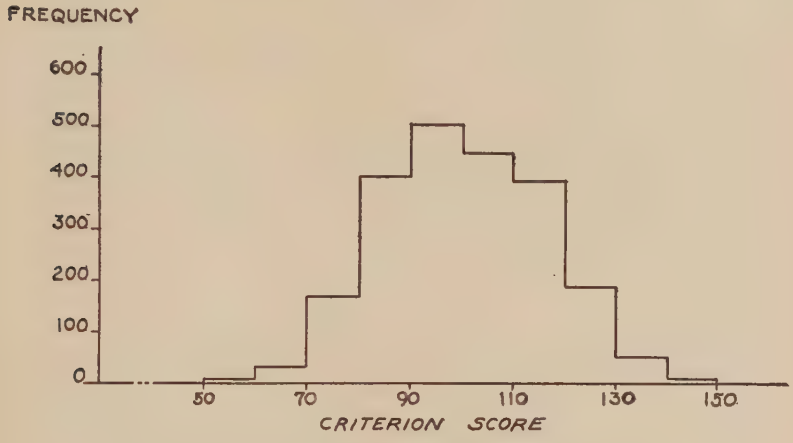


FIGURE 8
DISTRIBUTION OF CRITERION SCORES IN ENGLISH

In the determination of criterion scores in English there were included all students who had studied the subject eight terms and for whom there were recorded a school mark and a test score.[4] English is a constant in the program of studies each term for all students. In other subjects there were students who had studied the subjects for varying periods of time and with varying degrees of recency. For the purpose of this study the elimination of those variations in specific subject achievement which were due to the above causes, is essential. Accordingly in each subject other than English the criterion score group consisted of students who had studied the subject for the same number of terms and for whom the same period of time had elapsed since the subject was regularly on the students' schedules. The numbers of terms of study and elapsed time as well as the numbers of cases for the criterion score groups in the several subjects are reported in Table I.

[4] Certain test scores were discarded as indicated on p. 41.

No direct determination of the reliability of criterion scores was possible. The inclusion of average school marks probably reduced the reliabilities of criterion scores below those reported for the standardized tests in Table IV. In two subjects, algebra and plane geometry, it was possible to estimate approximately the lower limits of the reliability coefficients. Minimum reliability coefficients may be estimated for criterion scores in algebra and plane geometry by the use of the Spearman correction for attenuation and by assuming that the algebra and plane geometry criterion scores were equally reliable. The Spearman correction is as follows:

$$r_{\alpha\alpha} = \frac{r_{12}}{\sqrt{r_{11}} \cdot \sqrt{r_{211}}}$$

where $r_{\alpha\alpha}$ = relationship between "true" criterion scores in algebra and plane geometry, r_{12} = the obtained relationship between actual criterion scores in algebra with a reliability, r_{11}, and those in plane geometry with a reliability, r_{211}. $r_{\alpha\alpha}$ cannot be more than 1.00. Any value of less than 1.00 for this relationship indicates higher reliability coefficients for the actual criterion scores in these two subjects, since the reliability coefficients and the "true" relationships are inversely proportional. The obtained correlation coefficient for 62 students tested in both algebra and plane geometry was .712. Substituting assumed maximum value for $r_{\alpha\alpha}$ and the actual values for r_{12}, the Spearman attenuation correction formula becomes

$$1 = \frac{.712}{r_{11}} \qquad \text{whence}$$

$$r_{11} = .712$$

Accordingly, the minimum reliability coefficient for criterion scores in algebra or plane geometry is probably .712.

It is quite probable that the reliability of subject criterion scores lies between the reliabilities of the standardized tests and the school marks. Accepting .712 as the minimum reliability coefficient for criterion scores as established above, with an

upper limit probably set by the reliability of the test scores (median .91), conservatively estimated reliability of .80 to .85 seems justified for most criterion scores. This estimated reliability coefficient is higher than those reported for most school marks on pp. 46–49 and lower than those reported for most tests in Table IV, p. 47.

II. *The Measure of Mental Ability*

Scores in the Otis Self-Administering Test of Mental Ability, Higher Examination were the basic measures of mental ability. The problem of determining the most effective derived measure is presented. This measure should satisfy the following requirements:

1. It should be as accurate an index of average school achievement as it is possible to obtain from the given data.

2. It should be readily comparable with measures of mental ability used by other investigators.

3. Since one of the purposes of this study is the prediction of achievement, a measure should be used which remains relatively constant at different periods in the life of an individual.

At least six different measures of mental ability may be used. These are: (1) a raw test score, (2) a mental age, (3) a linear or sigma score derived from the distribution for the entire group tested and comparable with the linear score units obtained for the subject matter tests, (4) a T-score derived from the distribution of an unselected group of 12-year-olds, (5) an I. B. (Index of Brightness) derived from the age distributions of the author's norms and read from the chart furnished by the test manual, and (6) an IQ derived by the Otis method.[5] Raw scores and mental ages satisfy the first two requirements set but they do not remain relatively constant at different periods in the life of an individual. Linear units derived from the distribution

---

[5] IQ's derived by the Otis method are found by using either a table or chart furnished in the Manual of Directions for the test. This method offers the following advantages: (1) Mental growth is not assumed to stop abruptly at age 16, but slows down gradually from age 16 to 18. (2) IQ's are based upon position within the specific age groups rather than upon relationships to age norms.

and T-scores suffer the same limitation, and in addition they are not readily comparable with measures of mental ability used by other investigators. The index of brightness satisfies the first and third requirements but it is not readily comparable. The IQ meets the three requirements better than any other measure. In addition, at the chronological age level of senior high school graduation, 18 years, raw scores in the Otis test and IQ's are in equivalent units although measured from different zero points. Accordingly, whatever value is to be had from the use of raw scores is probably to be attained by the use of IQ's.[6] The most significant of these is the comparability of relationships determined in this investigation with those of other investigators who have used either raw scores (or their direct derivatives, mental ages) or IQ's. In the latter case, Otis IQ's may be looked upon as positions in a distribution of mental ability test scores at a given chronological age level rather than the ratio of the mental and chronological ages.

IQ's and achievement criteria were both determined in May 1928. Relationships between IQ's determined at the same time as the achievement criteria may not be indicative of the relationships which would exist if IQ's had been determined several years prior to the establishment of the achievement criteria. For predictive purposes relative constancy of the IQ must be assumed. However, this assumption is supported by considerable data. Studies summarized and reported by Otis,[7] Rugg,[8] Gates,[9] Dickson,[10] and Snow[11] show correlation coefficients of over .80 between mental measures repeated at various intervals up to

---

[6] Symonds favors the use of raw scores in mental ability tests for most purposes at the senior high school level. See Symonds, P. M.: *Measurement in Secondary Education*, 322.

[7] Otis, A. S.: *Otis Self-Administering Test of Mental Ability. Manual of Directions*, 12.

[8] Rugg, L. S.: "Retests and the Constancy of the IQ." *Journal of Educational Psychology*, 16: 341–343. May, 1925.

[9] Gates, A. J. and LaSalle, J.: "The Relative Predictive Values of Certain Intelligence Tests, Together with a Study of Effect of Educational Achievement upon Intelligence Test Scores." *Journal of Educational Psychology*, 14: 517–539. December, 1923.

[10] Dickson, V. E.: *Mental Tests and the Classroom Teacher*, 66.

[11] Snow, W. C.: *A Study of the Relationship between Mental Status and Classroom Achievement of Public School Pupils.*

3 years. In Rugg's study[12] the average difference in IQ over a period of 21 months was 3.1 points and the greatest change was 11 points.

A correlation coefficient less than 1.00 between two different IQ determinations, one of which has been obtained after the elapse of a period of time, limits the value of the relationships between IQ's and criterion scores reported in this study as predictive measures. This limitation, whether caused by the unreliability of the intelligence tests or by actual changes in the IQ would cause correlation coefficients between IQ's and criteria of specific subject success to be higher in this study than for measures determined some months or years apart. Other things being equal, then IQ criterion score correlation coefficients reported in this study may be more safely considered as maximum rather than minimum relationships for the purpose of prediction. However, most studies report correlation coefficients which are similarly limited. Accordingly, to facilitate comparisons, IQ–criterion score correlation coefficients reported in this study have not been corrected.

III. *The Measure of Relationship*

The measure most commonly used in investigations of the relationships between subject achievements and measures of mental ability is the Pearson product–moment correlation coefficient. Practically all investigators reported in Chapter I used this measure. Among other assumptions, however, this measure assumes rectilinearity of the relationship, *i. e.*, that the most probable values of one measure determined by values in the other will lie along a straight regression line, the slope of which is called the correlation coefficient.

In order to test the assumption of linearity for the regression of subject criterion on IQ, the correlation ratio, η (eta), was

---

[12] Rugg, *op. cit.*, 342.

determined as a measure of non-linear relationship[13] for each of five different subject criterion–IQ scatter-diagrams: English, Latin, physics, algebra, and American history. In four of the five, η was within one P. E. of the r, the Pearson product-moment coefficient. In the case of the IQ–English criterion relationship, η was but slightly more than 1 P. E. different. Applying Blakeman's[14] short test for rectilinear relationship to the data, it was found that in each of the five cases the relationship was rectilinear. Accordingly, the Pearson product-moment correlation coefficient was used as the measure of relationship in this investigation.

In addition to the advantage of comparability, the Pearson r's determined in this investigation may be used in multiple regression equations to add to the total validity of composite indices for the prediction of achievement. The use of the Pearson r is apparently justified.

The correlation coefficients were determined and calculated upon Otis correlation charts. Furnishing a convenient method of tabulating data, the scattergrams of the Otis charts provided opportunity for the calculation of the correlation ratios, means, and standard deviations used in this study.

---

[13] $$\eta_{yx} = \frac{\sqrt{\frac{\Sigma f_x (M_y - \overline{Y}_x)^2}{N}}}{\sigma_y}$$ where M_y = mean of the y measures; σ_y = S. D. of the y measures; $\overline{Y}_x$ = mean of a single array of y measures whose frequency is f; N = total number of cases in the distribution

Holzinger, *op. cit.*, 180.

[14] Holzinger, *op. cit.*, 183. This short test is applicable because $\eta^2 - r^2$ is small in comparison with r. Blakeman's formula is:

$$\sqrt{N} \cdot \sqrt{\eta^2 - r^2} < 4.047$$

CHAPTER V

Correlation Coefficients for Otis IQ's and Subject Criterion Scores

Table IX reports the raw correlation coefficients between Otis IQ's and each of the 16 subject-criterion scores thus affording measures of the existing relationships. These coefficients range from .575 for English to .123 for Spanish with .417 for solid geometry and .409 for economics in the middle of the ranked order. All coefficients are positive. All are significant to the extent of exceeding three times their probable errors with the exceptions of those for German and Spanish. Only two correlation coefficients indicate a degree of relationship between IQ's and subject-criterion scores adequate for the purpose of predicting one from the other within reasonable limits of error.[1] These are in English and civics.

As measures of the true relationships existing between the IQ and specific subject achievement, the raw correlation coefficients are inadequate. To determine true relationships accurately would require measures without errors. Such are not to be obtained in the field of educational measurements. The errors in measurement are due chiefly to lack of validity of the IQ and criterion score determinations. Contributing to the total error, lack of reliability of the measures affects the magnitude of the correlation coefficients. The effect[2] of variable errors in measurement probably has been to reduce the raw correlation coefficients reported in Table IX below those of accurate measures. Accordingly, the "true" relationships between IQ's

---

[1] Hull states that correlation coefficients below .45 or .50 are practically useless for differential prognosis and that those between .50 and.60 have some value. *Op. cit.*, 276.

[2] In *An Introduction to the Theory of Mental and Social Measurements*, Thorndike states (178) "the influence of chance inaccuracy in the measures to be related is *always to produce zero correlation.*" For a lucid description of the reasons for this effect see Hull, *op. cit.*, 237–243. Its correction is discussed by Spearman, C.: "The Proof and Measurement of Association between Two Things." *American Journal of Psychology*, 15: 72–101. January 1904.

TABLE IX

RAW CORRELATION COEFFICIENTS BETWEEN OTIS IQ'S AND SUBJECT CRITERION SCORES RANKED ACCORDING TO MAGNITUDE AND COMPARED WITH THE SAME COEFFICIENTS CORRECTED FOR DIFFERENCES IN THE VARIATION OF IQ'S

SUBJECT	GROUP TESTED		NO OF CASES[3]	RAW CORRELATION COEFFICIENT	P. E.$_r$	CORRECTED CORRELATION
	NUMBER TERMS INSTRUCTION	NO TERMS ELAPSED SINCE THE LAST SCHOOL MARK IN THE SUBJECT				
English	8	0	2209	.575	.012	.575
Civics	2	0	133	.552	.042	.563
Algebra	4	1	622	.488	.022	.492
American History	2	2	1742	.487	.013	.473
Ancient History	2	6	30	.476	.096	.561
Physics	2	0	178	.438	.048	.460
French	6	0	166	.422	.042	.443
Solid Geometry	1	3	51	.417	.077	.417
Economics	1	1	45	.409	.084	.416
Plane Geometry	2	3, 4	582	.400	.059	.414
General European History	2	4	110	.367	.056	.374
Trigonometry	1	0	98	.345	.059	.382
German	6	0	16	.333	.150	.413
Chemistry	2	0	122	.328	.055	.358
Latin	8	0	312	.315	.036	.295
Spanish	4	4	140	.123	.056	.142

and specific subject achievement are probably closer than thos indicated by the correlation coefficients reported. The ra correlation coefficients may be looked upon as probable minimu values of the "true" relationships.[4]

[3] The numbers of cases differ slightly from those reported in Table I becaus there were some students who had mental ability test scores and lacked achiev ment test scores and vice versa.

[4] This is only an approximate statement. Conceivably, constant errors ma have crept into the original data. In this case, the relationships would b different, probably, and the coefficients for the "true" relationships may b either higher or lower. The nature of such hypothetical errors is unknow and accordingly no correction can be applied.

For comparison of the differences in the degrees with which the various subject criterion scores are related to IQ's the correlation coefficients should be determined for groups which show exactly comparable variations in the distributions of scores. This was not the case with the raw correlation coefficients reported in Table IX. In some groups distribution variations both of IQ's and criterion scores were greater than in others. This may be seen in Table X reporting means and standard deviations of the distributions. For example, in Latin, the IQ variation for 312 students is expressed by a standard deviation of 10.3 about a mean IQ of 114.1. In English, for 2209 students, the IQ variation similarly measured was 9.6 about a mean of 107.3. These differences in variation invalidate direct comparison of r = .315, the correlation coefficient for IQ–Latin criterion scores, with r = .575, similarly determined for English. The equation used for the correction of the correlation coefficients for differences in variation of scores is given by Kelley:[5]

$$R_{12} = \frac{r_{12}\,\frac{\Sigma_1}{\sigma_1}}{\sqrt{1-r_{12}^2-r_{12}^2\left(\Sigma_1/\sigma_1\right)^2}}$$

where R_{12} is the correlation coefficient between IQ's and subject criterion scores for the extended range; r_{12} is a similar coefficient for the restricted range; Σ_1 is the standard deviation of IQ's for the extended range; σ_1 is a similar measure for the restricted range.

The application of this formula assumes that the common measure in all the correlation coefficients in Table IX, *i. e.*, the IQ, is the significant determiner of comparative variation. It further assumes that the variation of criterion scores is affected by variation of IQ's only in a corresponding manner. These assumptions are probably justified in view of the fact that differences in IQ variation can be accurately determined while real differences in the variation of criterion scores are not known, being measured on scales based upon achievement

[5] Kelley, T. L.: *Statistical Method*, 225.

scores for different groups. The assumption that the real criterion scores vary with the IQ distribution is more tenable than the assumption that the standard deviations of criterion scores for different subjects are strictly comparable. The formulae,

$$\frac{\Sigma_1}{\sigma_1} = \frac{\sqrt{1-r_{12}}}{\sqrt{1-R_{12}}} = \frac{\Sigma_2}{\sigma_2}$$

given by Holzinger,[6] Odell,[7] and Otis[8] are applicable only to correlation coefficients between two similar measures, *e. g.*, reliability coefficients.[9]

Since the group tested in English included all others, the variation of IQ's in the English–IQ distribution was taken as the base and all correlation coefficients were corrected to the variation within that distribution. In the column at the right of Table IX the corrected coefficients are reported. The corrected coefficients are within one P. E. of the raw coefficients for all subjects except American history where the corrected coefficient is lowered by an amount equal to slightly more than one P. E. Raw correlation coefficients were reduced in Latin and American history; in all other subjects, except solid geometry (and English) with unchanged coefficients, the correlation coefficients were raised.

Comparison of the various subject groups shows that in all social studies the corrected correlation coefficients are above .400, except that for general European history (.374). In mathematics all similar coefficients are above .400 except that for trigonometry (.382). Sciences are represented by physics and chemistry with corrected correlation coefficients of .460 and .358 respectively. The lowest corrected correlation coefficients are in foreign languages with .142 for Spanish and .295 for Latin, although .443 was obtained for French and .413 for German. With exceptions to be noted for individual subjects, the subject groups may be ranked from largest to smallest

---

[6] Holzinger, *op. cit.*, 253–254.
[7] Odell, *op. cit.*, 175.
[8] Otis, *op. cit.*, 235–237.
[9] See proofs, Kelley, *op. cit.*, 222. Holzinger, *op. cit.*, 253–254.

in order of magnitude of their corrected correlation coefficients as follows: English, social studies, mathematics, physical sciences, and foreign languages.

TABLE X

CRITERION SCORE AND I. Q. MEANS AND STANDARD DEVIATIONS FOR EACH SUBJECT GROUP

SUBJECT	NO. OF CASES	CRITERION SCORE		I. Q.	
		MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
English	2209	99.4	15.6	107.3	9.6
Civics	133	100.1	14.5	104.2	9.4
Algebra	622	100.8	15.6	111.8	9.5
American History	1742	100.1	15.6	108.2	10.0
Ancient History	30	102.0	13.2	101.3	7.7
Physics	178	100.6	16.4	110.5	9.0
French	166	100.4	15.1	110.1	9.0
Solid Geometry	51	105.5	12.8	113.4	9.6
Economics	45	100.6	13.6	97.9	9.4
Plane Geometry	582	100.1	15.3	111.0	9.2
General European History	110	101.8	14.8	103.8	9.4
Trigonometry	98	99.1	17.1	113.8	8.6
German	16	95.6	17.5	107.5	7.5
Chemistry	122	97.5	16.4	110.1	8.7
Latin	312	100.5	16.0	114.1	10.3
Spanish	140	101.6	15.4	104.0	8.2

For the purpose of investigating the relationships for these five major subject groups, the test items in the Otis Self-Administering Test of Mental Ability, Higher Examination, Form A, were placed in subject-matter[10] categories. For English, there were included items of vocabulary which are not commonly used in any of the other four major subject groups. Reading comprehension items were also included as "English" in the tabulation. The caption "Mathematics" included arithmetic problems, and items involving number, line, and space concepts. "Natural Sciences" included information items in biology, physics, astronomy, hygiene, and chemistry. "Social Sciences" included items in economics, history, logic, and social relations.

[10] The items of this analysis are reported in Appendix C, pp. 142-147.

No items were found which could be classified as "Foreign Languages". The tabulation is reported below:

English	28 Items
Mathematics	19 Items
Social Science	9 Items
Natural Science	9 Items
Foreign Languages	0 Items
Not Classified	10 Items
Total	75 Items

The highest correlation coefficient between Otis IQ's and any specific subject was that for English under which subject classification there is placed more than one-third of all the items in the test. The lowest similar correlation coefficients are for foreign languages in which category no items in the Otis test could be classified. Criterion scores in the social sciences of this investigation show somewhat closer relationships to Otis IQ's than do those of mathematics and natural sciences. However, the number of items in the Otis test, classified under the caption "Mathematics" is more than twice as great as those for "Social Science" or "Natural Science". Accordingly, the small differences between some of the IQ-subject criterion correlation coefficients could not have been anticipated as a result of studying Otis test content. The large differences in the correlation coefficients for IQ–English criterion score and the IQ–foreign language score might be due to the fact that more than one-third of the mental ability test content could be classified as "English" and none of it as "Foreign Language".

For five typical subjects, correlation ratios (etas) were determined and Blakeman's[11] short test for rectilinearity was applied. The correlation ratios are reported and compared with Pearson product-moment coefficients in Table XI. The greatest difference between the correlation ratio and the product-moment coefficient is .027. This may be considered as well within the limit of sampling error. The Blakeman test in each case indicated rectilinearity. Since there is no reason to

[11] See p. 66.

believe that other subjects should exhibit non-rectilinear relationships other than those of an accidental nature, it may be assumed that similar results would be obtained for the eleven subject distributions for which correlation ratios were not determined.

TABLE XI

CORRELATION RATIOS (ETAS) BETWEEN OTIS IQ'S AND SUBJECT CRITERION SCORES COMPARED WITH PEARSON PRODUCT-MOMENT CORRELATION COEFFICIENTS (r's)

SUBJECT	ETA	r
English	.561	.575
American History	.496	.487
Algebra	.491	.488
Physics	.465	.438
Latin	.298	.271

In order to study the relationships between IQ's and each of the two components of the subject criterion, *i. e.*, achievement test score and average school mark, Table XII has been provided. Correlation coefficients are reported[12] for the relationships between raw scores in the Otis test and raw scores in the subject achievement tests administered in the cooperative testing program which has furnished data for this study. Since IQ's are only transmuted raw scores at the chronological age level of 18 years, relationships determined for IQ's and similar relationships for raw scores in the Otis test are comparable.

For most subjects there is a higher degree of relationship exhibited by the data of Table XII between Otis scores and achievement test scores than there is between IQ's and average school marks. Thirteen of the sixteen coefficients for the relationship between Otis scores and achievement test scores are higher than comparable coefficients for IQ's and average school marks. In none of the remaining three cases does the coefficient for the correlation between IQ's and school marks exceed the comparable coefficient for Otis and achievement scores by an amount

[12] Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 109, *Subject-Matter Tests—Cooperative Program.*

equal to three times the probable error of the difference. In nine subjects the coefficients for Otis and achievement test scores are higher than comparable coefficients for IQ's and school marks by an amount exceeding three times the probable error of the difference. Twelve of the sixteen correlation coefficients expressing relationships between IQ's and the criterion scores (column (3), Table XII), lie in the ranges of values between comparable coefficients for raw score correlations (column (1), Table XII) and IQ school mark correlations (column (2), Table XII). In the remaining four cases, ancient history, economics, trigonometry, and German, the coefficients for IQ's and criterion scores were greater than the comparable coefficients reported in columns (1) and (2) for the separate constituents of the composite. Accordingly, the data of Table XII permit the following generalizations:

TABLE XII

CORRELATION COEFFICIENTS BETWEEN (1) SCORES IN OTIS TEST OF MENTAL ABILITY HIGHER EXAMINATION AND SUBJECT ACHIEVEMENT TEST SCORES, (2) OTIS IQ'S AND AVERAGE SCHOOL MARKS AND (3) OTIS IQ'S AND SUBJECT CRITERION SCORES

SUBJECT	(1) RAW OTIS SCORE AND RAW ACHIEVEMENT TEST SCORE [13]	(2) IQ AND AVERAGE SCHOOL MARK	(3) IQ AND CRITERION SCORES
English	.741 ± .006	.386 ± .012	.575 ± .010
Civics	.615 ± .027	.262 ± .055	.552 ± .042
Algebra	.670 ± .008	.338 ± .024	.488 ± .022
American History	.500 ± .010	.292 ± .019	.487 ± .013
Ancient History	.416 ± .080	.091 ± .122	.476 ± .096
Physics	.476 ± .025	.273 ± .047	.438 ± .048
French	.462 ± .021	.303 ± .048	.422 ± .042
Solid Geometry	.236 ± .068	.438 ± .079	.417 ± .077
Economics	.156 ± .058	.273 ± .093	.409 ± .084
Plane Geometry	.430 ± .017	.261 ± .026	.400 ± .024
General European History	.423 ± .043	.248 ± .060	.367 ± .056
Trigonometry	.342 ± .046	.239 ± .065	.345 ± .059
German	.280 ± .046	.167 ± .150	.333 ± .150
Chemistry	.331 ± .020	.315 ± .055	.328 ± .055
Latin	.300 ± .027	.333 ± .034	.315 ± .036
Spanish	.477 ± .024	.087 ± .056	.123 ± .056

[13] These coefficients are reported in Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 109, *Subject-Matter Tests—Cooperative Program.* Since IQ's are practically transmuted raw scores at the age level of this study, the data reported are comparable. Page references are given in Tables XIII to XVII inclusive.

(1) The degree of relationship shown by the correlation coefficients in most subjects (10 of 16) ranks from highest to lowest in magnitude in the order of raw score, IQ–criterion, and IQ–average school mark correlation.

(2) In two subjects, solid geometry and Latin, the relationships are exactly reversed in magnitude with IQ–average school mark coefficients highest and raw score coefficients lowest.

(3) In four subjects, ancient history, economics, trigonometry, and German, the relationships for the composite criterion are higher than for either constituent.

The relatively high coefficients for the relationships with raw test scores possibly may be explained by the following facts:

1. These scores have not been subjected to restriction in range as the result of averaging as in the case of the composite criterion scores.

2. Test scores attain generally higher degrees of reliability because they are not so dependent upon variable subjective judgments as are the school marks.

3. There are certain common factors in paper tests of both mental ability and specific subject achievement which tend to increase the correlation between the two measures.

Whether or not these facts explain the significantly higher relationships found between mental ability scores and achievement test scores, than between IQ's and average school marks, it is evident that the criterion of success must be considered in the evaluation of the relationship between IQ's and specific subject success. The composite criterion used in this study may be adequate for its purpose, but if specific subject success is to be measured by school marks alone in other studies, the correlation coefficients will probably be of a lower order of magnitude than those reported in Table IX. On the other hand, if other investigators use comprehensive test scores, the coefficients will probably be higher than those for the criterion scores of this study.

Tables XIII to XVII, inclusive, afford comparisons between the correlation coefficients determined in this study and those determined by other investigators. The subjects are grouped in each table under major fields: Table XIII, English; Table

XIV, social studies; Table XV, mathematics; Table XVI, physical sciences; Table XVII, foreign languages. For each subject the correlation coefficients are ranked according to magnitude, thus facilitating comparisons. Page references are given to Chapter II to permit further study.

TABLE XIII

CORRELATION COEFFICIENTS BETWEEN MEASURES OF MENTAL ABILITY AND ACHIEVEMENT IN ENGLISH, RANKED ACCORDING TO MAGNITUDE

CORRELATION COEFFICIENT	REPORTED[14]	CORRELATION COEFFICIENT	REPORTED[14]
.741	Bulletin 109, p. 4	.391	p. 26
.575	THIS STUDY, p. 68	.37	p. 23
.51	p. 23	.359	p. 29
.51	p. 27	.34	p. 27
.428	p. 26	.31	p. 27
.424	p. 26	.30	p. 27
.418	p. 29	.26	p. 27
.412	p. 26	.25	p. 27

It is difficult to be over-cautious in comparing the coefficients reported in Tables XIII to XVII inclusive. Correlation coefficients are probably the best available indices of relationship but they often do not show the degree of relationship accurately. Relationships which are essentially the same may yield quite different correlation coefficients. Differing relationships may be indicated by similar coefficients. Among the data necessary to make adequate comparisons, there should be (1) coefficients of the reliability of the measures of subject success and mental ability, (2) comparable measures of the variations of the scores for the groups studied, (3) assurance that the data were valid, accurately scaled, and that the computations were checked at every step, (4) assurance from inspection of the probable errors of the coefficients that they were not subject to wide variation due to chance errors of sampling. Unfortunately, few investigators can or do provide all the data necessary for adequate comparisons.

TABLE XIV

CORRELATION COEFFICIENTS BETWEEN MEASURES OF MENTAL ABILITY AND ACHIEVEMENT IN SOCIAL STUDIES, RANKED ACCORDING TO MAGNITUDE

A. AMERICAN HISTORY

CORRELATION COEFFICIENT	REPORTED[14]	CORRELATION COEFFICIENT	REPORTED[14]
.55	p. 23	.487	THIS STUDY, p. 68
.500	Bulletin 109, p. 9	.16	p. 27

B. ANCIENT HISTORY

.476	THIS STUDY, p. 68	.32	p. 27
.416	Bulletin 109, p. 47		

C. CIVICS

.615	Bulletin 109, p. 53	.11	p. 27
.552	THIS STUDY, p. 68		

D. ECONOMICS

.409	THIS STUDY, p. 68
.156	Bulletin 109, p. 51

E. GENERAL EUROPEAN HISTORY

.423	Bulletin 109, p. 49
.367	THIS STUDY, p. 68

However, comparisons may prove valuable in providing additional scope to the study of the relationships which exist between measures of mental ability and measures of achievement in specific subjects. Also, such comparisons provide an important method of checking the validity of the findings of this study. Of 16 correlation coefficients reported for this study in Tables XIII to XVII inclusive, 8 are above the middle coefficient in their respective distributions, 2 are middle coefficients, and 6 are below the middle coefficient. Accordingly, it may be said in generalizing that the correlation coefficients found in this study are slightly higher than those found by other investigators.

[14] Page references on which coefficients are cited are pp. 23 to 29, in this study. "Bulletin 109" refers to the Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 109, *Subject-Matter Tests—Cooperative Program.*

This effect may be due to the linear scaling procedure used in establishing subject criterion scores, and to the possibly greater reliability of the "criterion" scores of this study.

TABLE XV

CORRELATION COEFFICIENTS BETWEEN MEASURES OF MENTAL ABILITY AND ACHIEVEMENT IN MATHEMATICS, RANKED ACCORDING TO MAGNITUDE

A. ALGEBRA

CORRELATION COEFFICIENT	REPORTED[14]	CORRELATION COEFFICIENT	REPORTED[14]
.670	Bulletin 109, p. 37	.26	p. 28
.488	THIS STUDY, p. 68	.25	p. 28
.45	p. 23	.23	p. 28
.40	p. 27	.19	p. 28
.38	p. 28	.05	p. 27
.27	p. 23		

B. PLANE GEOMETRY

.57	p. 23	.400	THIS STUDY, p. 68
.45	p. 27	.40	p. 28
.430	Bulletin 109, p. 41		

C. SOLID GEOMETRY

.417	THIS STUDY, p. 68	.236	Bulletin 109, p. 43
.36	p. 27		

D. TRIGONOMETRY

.345	THIS STUDY, p. 68	.23	p. 27
.342	Bulletin 109, p. 45		

TABLE XVI

CORRELATION COEFFICIENTS BETWEEN MEASURES OF MENTAL ABILITY AND ACHIEVEMENT IN PHYSICAL SCIENCES, RANKED ACCORDING TO MAGNITUDE

A. CHEMISTRY

CORRELATION COEFFICIENT	REPORTED[14]	CORRELATION COEFFICIENT	REPORTED[14]
.50	p. 50	.331	Bulletin 109, p. 34
.407	p. 29	.330	THIS STUDY, p. 68
.38	p. 27	.270	p. 29

B. PHYSICS

.476	Bulletin 109, p. 31	.27	p. 27
.438	THIS STUDY, p. 68		

TABLE XVII

CORRELATION COEFFICIENTS BETWEEN MEASURES OF MENTAL ABILITY AND ACHIEVEMENT IN FOREIGN LANGUAGES, RANKED ACCORDING TO MAGNITUDE

A. FRENCH

CORRELATION COEFFICIENT	REPORTED[14]	CORRELATION COEFFICIENT	REPORTED[14]
.51	p. 25	.33	p. 24
.462	Bulletin 109, p. 13	.329	p. 26
.45	p. 24	.32	p. 28
.44	p. 25	.30	p. 23
.424	p. 25	.30	p. 24
.422	THIS STUDY, p. 68	.30	p. 24
.42	p. 24	.29	p. 24
.41	p. 24	.251	p. 29
.41	p. 25	.25	p. 24
.39	p. 25	.24	p. 25
.38	p. 25	.23	p. 25
.36	p. 25	.19	p. 25
.331	p. 26	.179	p. 29
		.092	p. 27

B. GERMAN

.333	THIS STUDY, p. 68	.280	Bulletin 109, p. 16

C. LATIN

.61	p. 24	.38	p. 27
.53	p. 24	.36	p. 24
.48	p. 24	.315	THIS STUDY, p. 68
.46	p. 23	.30	Bulletin 109, p. 19
.45	p. 27	.30	p. 27
.439	p. 23	.27	p. 27
.39	p. 28	.24	p. 27

D. SPANISH

.60	p. 24	.19	p. 24
.477	Bulletin 109, p. 22	.15	p. 27
.31	p. 24	.123	THIS STUDY, p. 68
.27	p. 28		

The correlation coefficients determined by the various investigators for the same subject groups (Tables XIII to XVII) vary in magnitude to about the same extent as do the raw correlation coefficients reported in this study for all subjects (Table IX). Thus, in Table XIII for English, the correlation coefficients vary from .25 to .741; in Table XIV social studies, from .16 to .615; in Table XV for mathematics, from .05 to

.670; in Table XVI for physical sciences from .27 to .50; in Table XVII for foreign languages, from .092 to .61. These variations may be compared with a range of raw coefficients reported for all subjects in this study in Table IX from .123 to .575. These wide variations imply not only inaccuracies of measurement but also lack of definition of specific subject achievement. Less variable correlation coefficients between measures of mental ability and specific subject achievement may be expected when achievement can be more accurately defined and measured.

CHAPTER VI

The Predictive Efficiency of the IQ For Specific Subject Achievement

The measure of predictive efficiency used in this study is Hull's index of predictive efficiency, E, as described on pp. 32, 33. This measure is probably more intelligible and significant than the correlation coefficient as an estimate of the value of the IQ in predicting achievement. The predictive efficiency index, E, is the per cent of perfect prediction attained by any forecasting device. If no errors exist in prediction, E = 100 per cent. If the error with the use of the predictive agent remains as great as prediction would be without it, E = 0 per cent. In this study, E is 100 per cent less the ratio of the error of criterion score prediction with the use of the IQ, to a similar error when prediction is entirely determined by chance. In other words, E is 100 per cent less the per cent of the total criterion score variation equal to the error in prediction of the criterion score, standard deviations being used for both measurements. In exact terms, E equals one minus the ratio of the standard error of prediction to the standard deviation of the distribution.[1]

The predictive efficiency indices for IQ's are given in Table XVIII. These vary from 0.8 per cent for Spanish to 18.2 per cent for English with 8.7 per cent for economics and 9.1 per cent for solid geometry in the middle of the ranked order. The IQ is apparently of negligible value as a sole index in the prediction of individual achievement, as measured in this study, for most subjects.

The per cents reported in Table XVIII are so low generally that the value of the IQ as a predictive index may be thought to be inferior to other indices of prediction. This is not the

---

[1] $E = 1 - \frac{\sigma\sqrt{1 - r^2}}{\sigma} = 1 - \sqrt{1 - r^2}$ The values of $\sqrt{1 - r^2}$ were obtained from: Miner, J. R.: *Tables of* $\sqrt{1 - r^2}$ *and* $1 - r^2$ *for Use in Partial Correlation and in Trigonometry.*

case. For example, the application of the predictive index formula to the highest correlation coefficient reported in Table I (.553) for the average school mark in plane geometry and the score in the Columbia Research Plane Geometry Test, shows that one predicts the other with an efficiency of only 16.8 per cent less error than pure chance placement. The constancy of the IQ is indicated by a correlation coefficient of .80[2] between repeated tests. This yields a predictive efficiency index, E, of 40 per cent less error than just a guess. A correlation coefficient of .921 is reported in Table IV between the scores in Otis Test of Mental Ability, Higher Examination, Form A, and scores in Form B over the grade range 7 to 12. This means that an individual's position in the distribution of scores for one form can be predicted from the position of a score in the other form of the test with an efficiency of 61.2 per cent.

TABLE XVIII

PER CENTS REDUCTION IN THE ERRORS OF PREDICTING SPECIFIC SUBJECT ACHIEVEMENTS USING OTIS IQ'S RATHER THAN PURE CHANCE PLACEMENT WITHIN THE ACTUAL SUBJECT GROUPS

SUBJECT	PER CENT REDUCTION IN ERROR (E)	SUBJECT	PER CENT REDUCTION IN ERROR (E)
English	18.2	Plane Geometry	8.3
Civics	16.6	General European History	7.0
Algebra	12.7	Trigonometry	6.1
American History	12.7	German	5.7
Ancient History	12.1	Chemistry	5.5
Physics	10.2	Latin	5.1
French	9.3	Spanish	0.8
Solid Geometry	9.1		
Economics	8.7		

Table XVIII reports predictive efficiency indices as they were actually determined from the raw correlation coefficients for each subject group. Accordingly, the data of Table XVIII are to be accepted as expressing the actual values of IQ's in predicting specific subject achievement in an existing situation rather than in an ideal situation in which the measures were

[2] See p. 64.

perfectly reliable and equal variation existed. The use of the raw correlation coefficients as indicated is justified by the fact that prediction is seldom, if ever, favored by the ideal situation described.

In Table XIX predictive efficiency indices are reported for the correlation coefficients corrected for differences in variation of IQ's. The resulting indices are significant in comparing the predictive values of the IQ for achievements in different subjects. They are not significant in comparing the findings of this study with those reported by other investigators unless those findings are similarly corrected. Table XIX shows that the indices corrected for differences in IQ variation range from

TABLE XIX

PER CENTS REDUCTION IN THE ERRORS OF PREDICTING SPECIFIC SUBJECT ACHIEVEMENTS USING OTIS IQ'S RATHER THAN PURE CHANCE PLACEMENT WITHIN HYPOTHETICAL SUBJECT GROUPS OF THE SAME IQ VARIATION AS THAT OF THE TOTAL GROUP

SUBJECT	PER CENT REDUCTION IN ERROR (E)	SUBJECT	PER CENT REDUCTION IN ERROR (E)
English	18.2	Economics	9.2
Civics	17.3	German	8.9
Ancient History	17.2	Plane Geometry	8.0
Algebra	13.0	Trigonometry	7.6
American History	11.9	General European History	7.3
Physics	11.2	Chemistry	6.6
French	10.3	Latin	3.4
Solid Geometry	9.2	Spanish	1.0

1.0 for Spanish to 18.2 for English with 9.2 for solid geometry and economics in the middle of the ranked order. The differences between predictive indices reported for the same subjects in Tables XVIII and XIX are small enough to show that differences in variation of IQ's do not account for the relatively low values of the IQ in the prediction of specific subject achievement.

The fact that the IQ has apparently but little value for the prediction of specific subject achievement as measured in this study only serves to emphasize the fact that effective prediction of school achievement is difficult to obtain through the use of a readily obtained index. Allen[3] found that the addition of scores in successive tests to a prognostic total score predicted success in Latin better than any single test score. He states further,

> "Terman Group Test of Mental Ability B was then tried and was found to add only .004 to the combination. It was therefore ellmininated.[4]"

The retained tests were Briggs' Analogies, Word Knowledge and Interpolation.[5]

Generalizing for this study, the IQ predicts the specific subject achievement of average individuals with variable errors which are from 1 to 18 per cent less than those of pure chance placement. Whether or not these low per cent reductions in error are of enough value to warrant the use of IQ determinations in the prediction of specific subject achievement is an administrative problem. Any reduction in error is desirable. However, it is important to realize that the purposes for which so-called tests of mental ability are used are seriously limited by the lack of their predictive value for specific subject achievement. Some of these purposes are:

1. The setting of individual achievement standards.
2. The use of the IQ as a measurement of individual aptitude for achievement in each of a number of specific subjects.
3. The grouping of pupils in specific subjects upon the basis of probable homogeneity of achievement indicated by a narrow range of IQ's, and

---

[3] Allen, *op. cit.*, 19.
[4] Allen, *op. cit.*, 20.
[5] A number of test scores are most effectively combined into a single predictive index by the use of the multiple regression equation. A discussion of the use of the multiple regression equation for prediction is given in Hull, *op. cit.*, 457–466. The multiple regression equation involves the use of the numerators of the Pearson product-moment equations for all correlations and intercorrelations of the predictive tests and the criterion.

4. The evaluation of the achievements of individuals by comparison of measures of their several achievements with measures of mental ability.

These purposes are probably important enough to warrant continuing the search for means of supplementing and improving the existing predictive devices.

CHAPTER VII

The Relationships Between IQ's and Subject Criterion Scores in Different Parts of the IQ Range

Two supplemental methods have been used in this study to determine differences in the relationships of IQ's to subject achievement in the different parts of the IQ range. The first method involved the calculation of the correlation coefficients between IQ's and subject criterion scores for IQ's of 110 or over. The process was repeated for IQ's lower than 110. The resulting coefficients were corrected for differences in the variation of IQ's and compared. The second method involved the determination in the scatter diagrams of the means of the successive arrays of criterion scores at each IQ level. The line of means so determined was plotted and compared with the slope of the regression line for the entire distribution. These more intensive analyses were undertaken for five subjects, which were selected to represent each of five major academic fields. The specific subject selected afforded the largest numbers of cases in each major field. These subjects were: English, algebra, American history, physics, and Latin, representing major fields of English, mathematics, social studies, natural science, and foreign language.

I. *Correlation Coefficients Between IQ's and Subject Criterion Scores for IQ's of 110 or over and IQ's lower than 110*

Each scatter diagram provided by the Otis Correlation chart was divided by a line drawn between the two IQ class intervals which were nearest the mean IQ of the distribution. In four of the five scatter diagrams this line divided the distribution at 110 IQ. In the Latin distribution the line fell at 116 IQ.

The ten raw correlation coefficients for the truncated distributions are reported in Table XX (column 2). Each raw correlation coefficient for the truncated distribution is smaller than the corresponding coefficient for the total distribution reported in Table IX. This finding is to be expected as a result

of the reduction in the variation of IQ's and criterion scores. The standard deviations of the truncated distributions of IQ's are reported in Table XX (column 3). In each case these measures of spread are reduced to slightly more than one-half the magnitude of the corresponding standard deviations reported in Table X. For example, the standard deviation for the IQ's of 110 or over in the IQ–English criterion distribution is 5.24 IQ units (Table XX, column 3). The standard deviation for the entire distribution is 9.60 IQ units (Table X).

The reduction in criterion score variation is much less than the reduction in IQ variation because the division of the distribution was on the basis of the IQ's rather than upon the basis of criterion scores. Comparison can be had by referring to Tables X and XX, column 4. Variation of criterion scores for each of the two parts of the distribution is less than the variation of the total distribution in each subject, except Latin. The upper IQ group in that subject provided a wider comparative spread of achievement than in the total distribution. This is accounted for by the presence of some low criterion scores attained by relatively high IQ students.

On account of the differences in the variation of IQ's for each truncated distribution, comparison between raw correlation coefficients for the upper and lower parts of the IQ distribution in any given subject has but little meaning. Table XX (column 5) reports correlation coefficients for each part of the distribution corrected[2] to the same variation of IQ's as that of the total distribution in each subject. None of these corrected coefficients differ from the raw coefficients reported in Table IX for the entire distributions by more than three times the probable error of the uncorrected coefficients (Table XX, column 6).

The corrected coefficients in Table XX for the upper and lower parts of each IQ distribution in each subject may be compared with each other. The probable error of the raw correlation coefficient is also reported to furnish an estimate of the reliability of the difference between two coefficients in each subject. In English, algebra, and physics the corrected correlation co-

[2] The method used was exactly the same as that described on pp. 69–70.

TABLE XX

CORRELATION COEFFICIENTS BETWEEN IQ'S AND SUBJECT CRITERION SCORES FOR UPPER AND LOWER IQ GROUPS: RAW AND IQ VARIATION CORRECTED COEFFICENTS[1]

		1	2	3	4	5	6
		N	RAW CORRELATION COEFFICIENT	STANDARD DEVIATION OF IQ'S	STANDARD DEVIATION OF CRITERION SCORES	CORRECTED CORRELATION COEFFICIENTS	P. E. RAW CORRELATION COEFFICIENTS
English	IQ's of 110 or over	945	.354	5.24	14.05	.572	.019
	IQ's under 110	1264	.354	5.86	13.58	.527	.017
Algebra	IQ's of 110 or over	353	.403	5.62	14.80	.605	.030
	IQ's under 110	269	.255	4.99	13.98	.451	.038
American History	IQ's of 110 or over	736	.186	3.77	14.48	.447	.024
	IQ's under 110	1006	.305	6.10	14.40	.465	.019
Physics	IQ's of 110 or over	96	.390	5.85	16.30	.567	.058
	IQ's under 110	82	.242	5.45	15.30	.379	.070
Latin	IQ's of 116 or over	126	.119	4.15	17.15	.282	.058
	IQ's under 116	186	.258	5.95	14.88	.416	.046

[1] Correlation coefficients are corrected to the variation of IQ's of the total distribution in each subject by the method described on pp. 69–70. This method does not correct completely for the mutilation of the distributions but is probably accurate enough if comparison is restricted to the two parts of the same subject distribution. No comparisons with other correlation coefficients are valid. See Holzinger, *op. cit.*, 172.

efficients for the upper IQ groups are greater than those for the lower groups. In two of these subjects, algebra and physics, the differences approximate three times the probable error of the difference of the raw coefficients.[3] In American history and Latin, the corrected coefficients are smaller in the upper groups. However, the differences for these two subjects are less than three times the probable error of the difference of the raw coefficients. Accordingly, all differences in relationships between IQ's and subject criterion scores for the upper and lower parts of the IQ distributions are possibly due to chance with the probable exception of the differences for algebra and physics. In these subjects the relationships between Otis IQ's and subject criterion scores are probably closer for the higher IQ groups.

Differences in relationship and, therefore, in the value of IQ determinations for the prediction of specific subject achievement are not particularly evident when the distribution of IQ's is divided into two approximately equal parts. This finding is further substantiated by the fact that the distributions in the scatter diagrams are essentially linear.[4] If there were important differences in the relationships for the upper and lower parts of the IQ range, the total relationship would be curvilinear and more accurately measured by the correlation ratio, eta.

II. *Criterion Score Means of the Successive Arrays at Different IQ Levels Compared with the Regression Lines for Total Distributions*

For the purpose of investigating differences in relationships not revealed by separate study of the two approximate halves of the IQ distributions, recourse was had to a second method. This is illustrated by application to distributions in English. Results are reported also for the four other subjects studied above: algebra, American history, physics, and Latin.

---

[3] The probable error of the difference of uncorrelated measures (r's are measures) is given by the equation P.E. dif $= \sqrt{(PE_1)^2 + (PE_2)^2}$ in which P.E. dif = probable error of the difference. P. E. $_1$ and P. E. $_2$ are the probable errors of the correlation coefficients for the upper and lower IQ groups respectively. Additional errors due to the use of the correction for differences in heterogeneity were not considered.

[4] See pp. 65, 66.

TABLE XXI

Correlation Distribution for I. Q.'s and Criterion Scores in English

CRITERION SCORE INTERVALS	NUMBER OF CASES											CRITERION SCORE FREQUENCIES
149 140								1			2	3
139 130						1	8	7	16	14	7	53
129 120				2	10	17	45	46	47	16	7	190
119 110	1		6	10	45	81	95	83	47	24	3	395
109 100		2	10	42	72	108	103	71	35	13	1	457
99 90	2	8	27	79	95	126	94	51	20	6		508

89 80	6	12	42	73	113	89	44	16	5	1		401
79 70	4	17	28	50	30	25	10	3	1			168
69 60		3	7	10	4	2	3					29
59 50			2	1	2							5
I. Q. Intervals	80 84	85 89	90 94	95 99	100 104	105 109	110 114	115 119	120 124	125 129	130 134	
I. Q. Frequencies	13	42	122	267	371	449	402	278	171	74	20	2209
Mean Criterion Scores of Successive I. Q. Distributions	85.8	82.6	86.4	89.8	95.0	99.0	104.6	108.9	114.1	117.2	128.0	
Mid-Point of I. Q. Intervals	82.5	87.5	92.5	97.5	102.5	107.5	112.5	117.5	122.5	127.5	132.5	

$r = .575 \pm .012$ Means { Criterion = 99.4; I. Q. = 107.3 } Standard Deviations { Criterion = 15.6; I. Q. = 9.6 }

Table XXI is the scatter diagram for IQ's and criterion scores in English. The means of the vertical distributions of criterion scores (columns) are reported in the next to the bottom row of the table. The bottom row reports the midpoints of the successive IQ distribution groups. Each of these measures, *i. e.*, the criterion score mean and the midpoint of the IQ interval, was converted to a standard deviation position in the respective total distribution. This was done by finding the difference between each of these scores and the mean of the same total distribution and dividing the difference by the standard deviation of the distribution. Thus, for the IQ interval 80-84, a criterion score mean of 85.8 is reported. This is 13.6 criterion score units below 99.4, the mean criterion score in the total distribution. This difference, divided by 15.6, the standard deviation of the distribution, locates a position 0.87 standard deviations below the mean. The midpoint of the IQ interval, 82.5, was similarly determined at 2.60 standard deviations below 107.3, the mean IQ in the total distribution. The point 0.87 standard deviations below the mean criterion score and 2.60 standard deviations below the mean IQ was then located in Figure 9, which is laid off in standard deviation units. Similar points, determined for each of the columns in Table XXI, were located in Figure 9 by circles with central dots. These points, connected by a solid and slightly curved line, are labelled line of the means of arrays in the figure. The average slope of this line is indicated by the dotted "Regression line" of criterion scores on IQ's. The regression line slopes from the horizontal at an angle whose tangent is .575. This is the correlation coefficient for IQ's and English criterion scores, since both IQ's on the horizontal axis and English criterion scores on the vertical axis are measured in equal standard deviation units. "Perfect correlation" is indicated by the slope of the solid line at an angle of 45°, *i. e.*, an angle whose tangent is 1.0. Accordingly, in Figure 9, changes in the slope of the line of the means of arrays may be readily noted at different parts of the distribution.

With the exception of the points for the lowest and highest IQ arrays, the line of the means of arrays in Figure 9 approximates the regression line with considerable fidelity. Reference

to Table XXI shows that the lowest IQ array includes IQ's below 85 and that the highest includes IQ's of 130 or over. Students with IQ's below 85 and those with IQ's of 130 or over attained mean English criterion scores higher than would be estimated from the total slope of the dotted regression line. In the case of the group of IQ's below 85, IQ's are not so closely related to English criterion scores as are criterion scores and IQ's above 130.

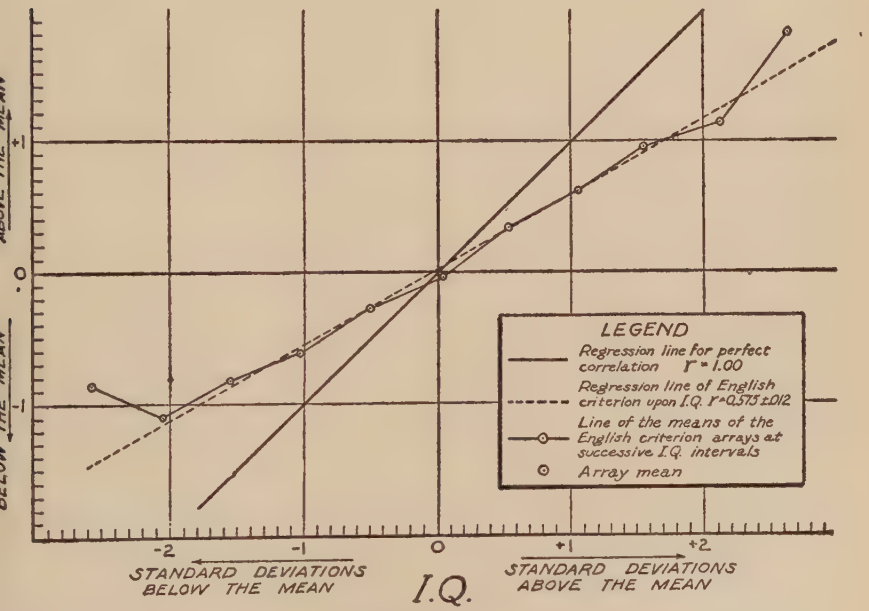


FIGURE 9

COMPARISON OF THE MEANS OF THE ENGLISH CRITERION SCORE ARRAYS WITH THE REGRESSION OF ENGLISH CRITERION UPON I.Q. (N = 2209)

Figures 10, 11, 12, and 13 have been prepared for algebra, American history, physics, and Latin respectively, in a manner similar to that described for English in Figure 9. Figures for those subjects in which the larger numbers of cases are reported, namely English, algebra, and American history show that the lines of the means of the arrays are smoother and approximate

more nearly the regression lines than in those subjects where the number of cases is smaller, *i. e.*, physics and Latin. In every case the criterion score mean for the highest IQ group (130 and above) is higher than that predicted by the slope of the regression line for the entire distribution. In three of five subjects, English, algebra, and American history, the criterion score means for the lowest IQ group (below 85) is higher than the points indicated by the slope of the regression line for the

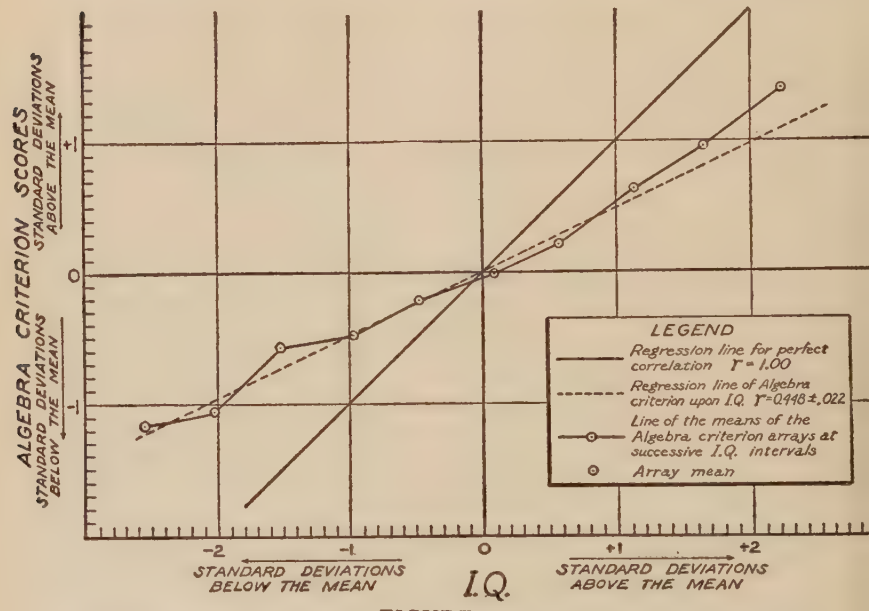


FIGURE 10
COMPARISON OF THE MEANS OF THE ALGEBRA CRITERION SCORE ARRAYS WITH THE REGRESSION OF ALGEBRA CRITERION UPON I.Q.
(N = 622)

entire distributions. The two exceptions in the case of IQ's below 85 are caused by a single criterion score each, one in physics below the point indicated by the regression line, and one in Latin on the regression line. With the exception of these two points determined by a single score in each case, IQ arrays of 130 and above and arrays below 85 IQ have subject criterion means above those indicated as most likely by the

slope of the respective regression lines. IQ arrays of 130 and above were composed of 20 cases in English, 14 in algebra, 16 in American history, 3 in physics, and 6 in Latin. Similarly, the three subjects in which the mean criterion scores of IQ arrays below 85 were higher than the scores predicted by the slope of the regression line, were English, algebra, and American history with 13, 4, and 11 cases respectively in the lowest IQ

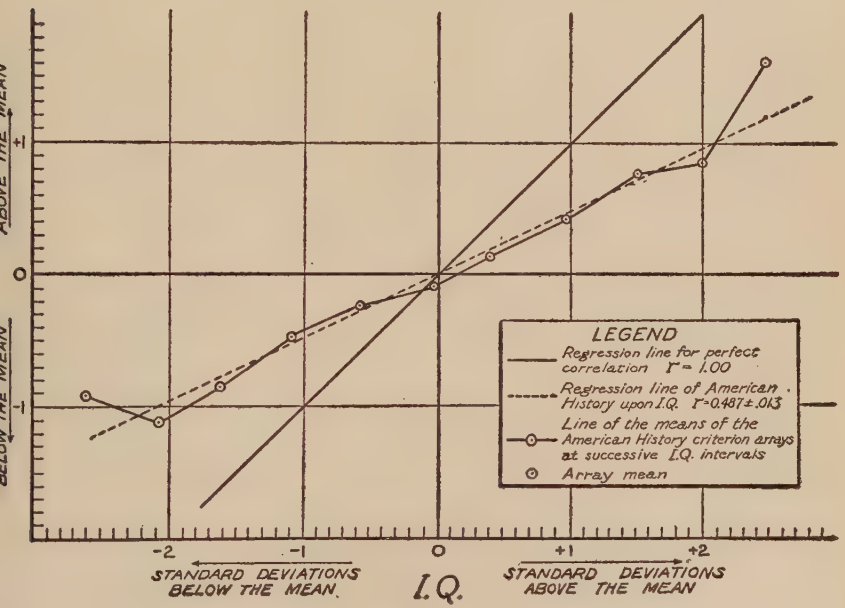


FIGURE 11

COMPARISON OF THE MEANS OF THE AMERICAN HISTORY CRITERION SCORE ARRAYS WITH THE REGRESSION OF AMERICAN HISTORY CRITERION UPON I.Q.

(N = 1742)

interval. Accordingly, the finding of the criterion scores higher than predicted by the regression line for the extremes of the IQ range is based upon results for 59 cases for high IQ's and 28 cases for low IQ's. The two individual score entries noted as exceptions can hardly be considered as invalidating the finding.

It is possible that the IQ furnishes a more accurate prediction of specific subject achievement for the highest group of IQ's than for any other group. Also, it is possible that qualities other than those which are measured in a group test of mental ability are of more importance than the IQ in determining the subject achievements of students with IQ's below 85 and who

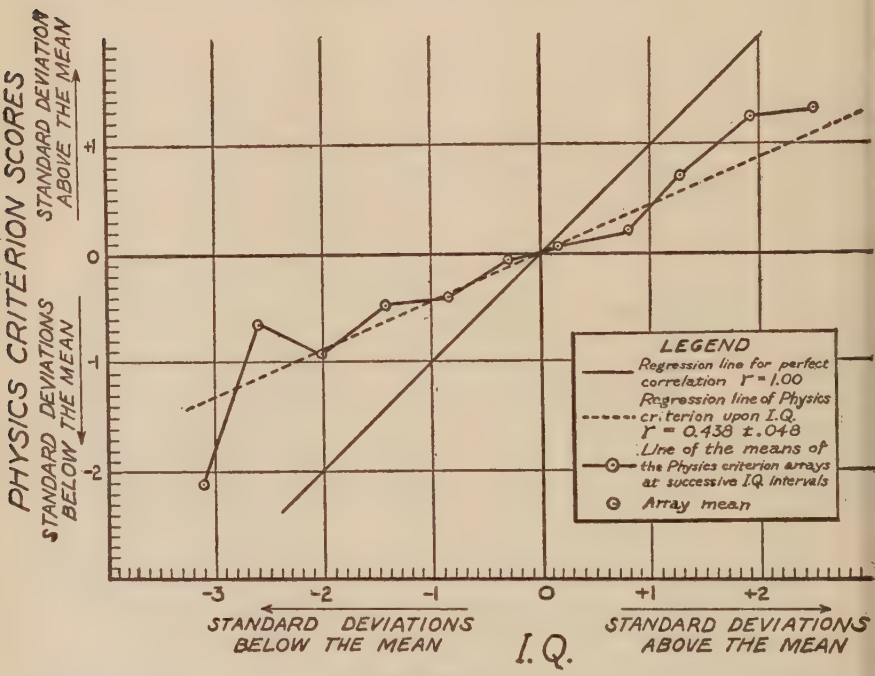


FIGURE 12

COMPARISON OF THE MEANS OF THE PHYSICS CRITERION SCORE ARRAYS WITH THE REGRESSION OF PHYSICS CRITERION UPON I. Q.
(N = 178)

persist in high school to the point of graduation. The existence of these other qualities, and consideration of the fact that individuals with low IQ's can and do exceed the achievements of their more "capable" fellows in academic subject fields should cause hesitancy in making any recommendation that students

with low IQ's should, without other consideration, be advised that they cannot succeed in the academic work of senior high schools. On the other hand, this study shows that students with IQ's over 130 attain specific subject achievement ranks more nearly comparable with the promise afforded by their IQ ranks.

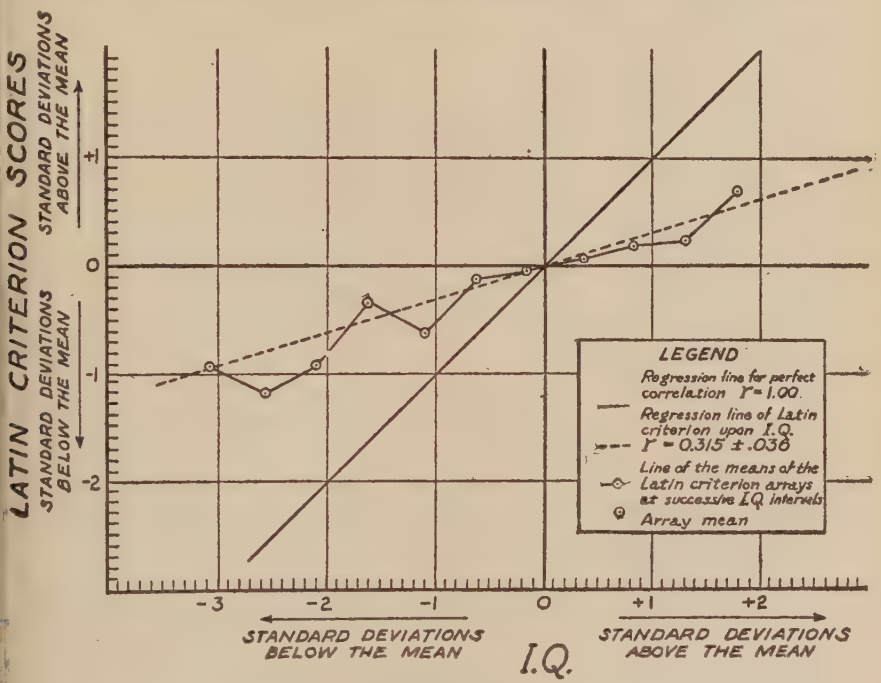


FIGURE 13

COMPARISON OF THE MEANS OF THE LATIN CRITERION SCORE ARRAYS WITH THE REGRESSION OF LATIN CRITERION UPON I.Q.
(N = 312)

CHAPTER VIII

The Extent of Individual Variation in Specific Subject Achievement and the Limitations Placed Upon the Prediction of Specific Subject Achievement by the Use of Any Index of General Achievement

I. *Preparing Equated Scales of Specific Subject Achievement*

By "individual variation in specific subject achievement" is meant the extent to which measured achievement in each specific subject for an individual student varies with respect to the level of his mean achievement score in all subjects. A common method of observing such variation is to examine the irregularity of the line in an achievement profile chart.[1]

In order to prepare such profile charts it is desirable to have equated achievement scales so that a given horizontal line on the chart cuts the separate scales at equivalent positions in the distributions. Equated scales of achievements are usually obtained by the application of scaling techniques to the distributions of scores for the same pupils in several subjects. As a hypothetical case, if 1000 pupils were tested at the same time in sixteen different subjects, a distribution of 1000 scores in each of the sixteen tests would furnish data for the preparation of an equated scale. A score in a given test could be placed at a specific point in the distribution of the 1000 scores for that test. The construction of 16 equivalent scales would then be simply a matter of developing linear or sigma scores by a method similar to that described in Chapter IV (pp. 52–56).

Unfortunately for the present purpose, students in senior high schools select programs of studies which differ widely from each other. Of the 16 subjects concerned in this investigation only two, American history and English, were constants and therefore required of all students. The testing program involved the participation of each student in five tests: one in American

[1] See Figures 14 and 15, pp. 102 and 103 or any form of the Stanford Achievement Test or the New Stanford Achievement Test.

history, one in English, and three in selected subjects. Accordingly, a wide variety of subject combinations resulted for the different individual students. The groups selected in each subject for the present study were those who had been instructed for the same numbers of terms and with the same degrees of recency. This fact introduced additional complexity in the derivation of equated scales.

As a result of these limitations of the data there was necessitated the development of a technique which would provide subject criterion scales equated as if all students had been tested in all subjects and as if these complete but hypothetical distributions had been used for scaling criterion scores. The assumptions underlying this technique undoubtedly do not add to the reliability of the findings, but it is believed and demonstrated that the errors introduced are of negligible proportions. The alternative of setting up conditions which probably would be extremely artificial with respect to the determination of subject achievement in the high school is less desirable than the complexity of technique involved in the interpretation of obtained results.

Uncorrected criterion scores were available being obtained by the method described on pages 52–61. The problem, then, was to so correct these subject criterion scores in all subjects that they would approximate positions in the hypothetical distribution of criterion scores which would be obtained if criterion scores existed for all students in all subjects. The distribution of criterion scores which includes the scores of approximately all students was that of English. Comparisons between the criterion scores in each subject and those of English were possible through the medium of comparable IQ distributions. For example, the mean IQ of the total group of 2209 students in the English criterion group is 107.3. In the Latin group of 312 students the mean IQ is 114.1. These data are reported in Table X. From the fact that there is a positive correlation of .315 (Table IX) between IQ's and Latin criterion scores, it may be assumed that if all of the 2209 students in the English subject group (mean IQ = 107.3) were also measured for Latin achievement, their mean achievement score in Latin would be

on a somewhat lower scale than that attained by the group of 312 students whose mean IQ was 114.1. The degree to which the mean Latin achievement score would be lowered if 2209 students were included in the Latin criterion group, may be estimated by the use of the regression equation:[2]

$$y = r \frac{\sigma_y}{\sigma_x} x$$

where y = the number of Latin criterion score units, the mean criterion score of pupils with IQ of 107.3 is less than the mean criterion score of all (312) students in the Latin criterion group; r = correlation coefficient between IQ's and Latin criterion scores, or .315 (Table IX); σ_y = the standard deviation of the Latin criterion scores, or 16.0 (Table X); σ_x = the standard deviation of the IQ's of 312 students in the Latin criterion group, or 10.3 (Table X).

Accordingly, students with a mean IQ of 107.3 may be expected to obtain a mean criterion score 3 units below the mean of the 312 students in the Latin criterion group. To bring Latin criterion scores to equality with the English criterion scores, Latin criterion scores should be corrected by the addition of 3 units. In a manner similar to that described for Latin criterion scores, corrections were determined for each of the criterion score scales. These corrections are given in Table XXII.

TABLE XXII

NUMBERS OF UNITS ADDED TO (+) OR SUBTRACTED FROM (−) THE CRITERION SCORES IN EACH SUBJECT TO PRODUCE COMPARABLE SCALES AT THE SAME LEVELS

(Determined from the IQ-Criterion Score Regression Equations)

SUBJECT	CORRECTION	SUBJECT	CORRECTION
Civics	−3	Economics	−7
Algebra	+4	Plane Geometry	+2
American History	+1	General European History	−2
Ancient History	−5	Trigonometry	+4
Physics	+3	German	0
French	+2	Chemistry	+2
Solid Geometry	+3	Latin	+3
		Spanish	−1

[2] Odell, C. W.: *Educational Statistics*, 192.

Most corrections are relatively slight, 11 of 16 being three criterion score units or less. Since most criterion score distributions have standard deviations approximating 15 units, the correction in most cases is 0.2 standard deviation or less.

The basic assumption of this technique is that the mean subject criterion score for groups can be predicted by the mean IQ of the group. This assumption is fundamentally different from an assumption that individual criterion scores can be predicted from individual IQ determinations. Data cited in Chapter VI indicate that the accurate prediction of individual criterion scores through the use of the IQ's in this study is not possible. However, the prediction of mean criterion scores is subject to much less error. This may be seen by examination of the actual means of the criterion score arrays in Figures 9, 10, 11, 12, and 13 at the two IQ intervals on either side of the means of the total distributions. The maximum divergence of the array means is approximately 0.1 standard deviation from the positions indicated by the regression lines. The numbers of cases involved in these arrays are less than those in the total distributions for the several subject scales to which the corrections are applied. It is quite probable that the error introduced by the correction of the scales is less than two criterion score units for most subjects. This amount[3] of error may be considered almost negligible in view of the magnitudes of the variations of individual scores measured by this method and reported below.

II. *Profile Charts*

There were 183 individuals for whom uncorrected criterion scores had been prepared in each of 5 different subjects. This number represents those pupils who were enrolled in each of five

---

[3] The probable errors of the means of the criterion score distributions are significant measures of the variation of these means from the positions indicated by the regression lines. Accordingly, the P. E.mean may be used as a measure of the error introduced by this technique $\left(\text{P. E.}_{\text{mean}} = .6745\frac{\sigma}{\sqrt{N}}\right)$

These errors are less than 2 criterion score units in all scales used in this part of the investigation, except in the scale for German (4 units).

tested subjects for the same periods of time as were those students who constituted the criterion groups. Profile charts are provided in Figures 14 and 15 upon which are plotted the criterion scores for two individuals. The scales used in these charts are those which have been corrected as described above. The wide variation shown in Figure 14 may be contrasted with the relative uniformity of achievement shown in Figure 15. In Figure 14, the divergence of the Latin and algebra scores from the mean criterion score is a measure of the degree to which one individual's achievements in specific subjects depart from his average achievement. The difference between the scores in these two subjects for the same individual is 50 criterion score units. His score

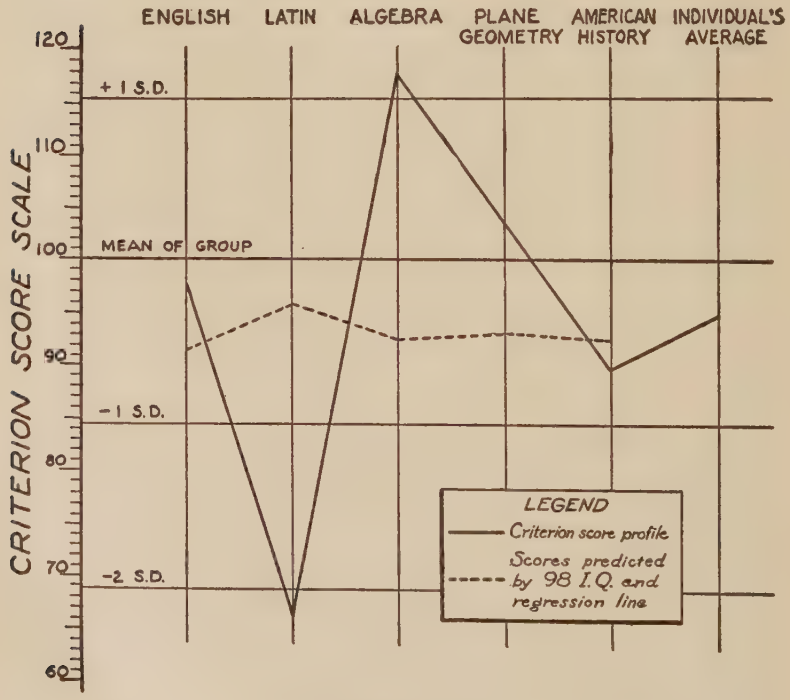


FIGURE 14
INDIVIDUAL PROFILE CHART SHOWING WIDELY VARIABLE CRITERION SCORES

in algebra is among the highest 16 per cent while his score in Latin is among the lowest 2 per cent of the respective distributions. This fact illustrates the difficulty of predicting specific subject achievement by any measure of general achievement for some individuals.

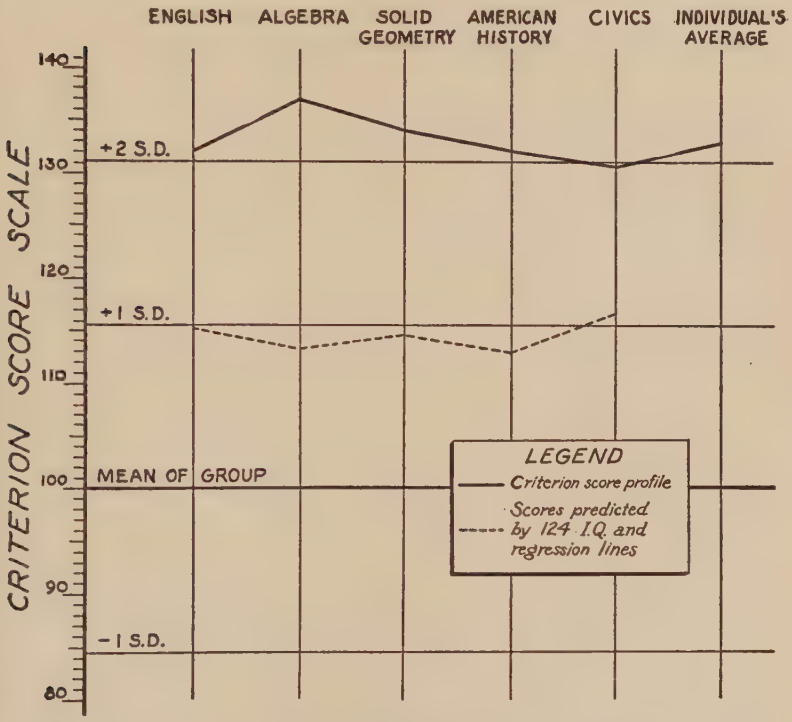


FIGURE 15

INDIVIDUAL PROFILE CHART SHOWING CLOSELY SIMILAR CRITERION SCORES

On the other hand, the uniform level of the criterion scores given in Figure 15 indicates that for some individuals an accurate index of general achievement may serve for predicting specific subject achievement. In order to study whether the pupils with high mean criterion scores, or those with low scores, exhibited the greater average degrees of variation in achievement, the stand-

ard deviations for the 183 distributions represented in the individual profile charts were correlated with the 183 individual means. In other words, measures of variation in achievement were correlated with measures of average achievement. The obtained coefficient was $-.034 \pm .050$ and therefore no relationship is indicated. A similar finding of no significant difference in "specialization of abilities" for gifted children in comparison with normal children is reported by De Voss.[4]

III. *The Correlations Between Average and Specific Achievement*

While individuals may vary greatly in the degree to which their achievements in specific subjects depart from their averages, as illustrated in Figures 14 and 15, it is desirable to know the average degree of such variation in achievement. This is found by determining the variation of the criterion scores in the separate subjects about the mean criterion score for each individual. In other words, those subjects in which many criterion scores departed significantly from the individual mean criterion scores were subjects of greater individual variation in achievement than those in which but few criterion scores differed significantly from the mean. The closer the relationship between the mean criterion scores and the scores in any one subject, the smaller the degree of variation in that subject.

A convenient and significant method of determining this relationship seemed to be the use of the correlation coefficient between mean criterion scores and each of the subject criterion scores. There were 183 students for whom it was possible to determine mean criterion scores in 5 subjects. The criterion scores for these students were scattered among 13 subjects. These subjects are listed in Table XXIII, which gives the correlation coefficients between mean criterion scores and each subject criterion score. Since the heterogeneity of the criterion score distributions of these groups of students differed from each other and from that of the total group, correlation coefficients corrected for differences in dispersion of mean criterion scores

---

[4] De Voss, J. C.: "Specialization of the Abilities of Gifted Children." Ch. XII in Terman, L. M.: *Genetic Studies of Genius*, Vol. I, 307–362.

also are given for comparison. The method of correction is similar to that reported for Table IX, p. 68. The subjects are listed in the order given in Table IX with the exception of ancient history, general European history, and economics for which the numbers of cases were so small that no significance could be attached to the correlation coefficients. This arrangement facilitates comparison with the coefficients reported in Table IX.

TABLE XXIII

RAW CORRELATION COEFFICIENTS BETWEEN THE MEAN CRITERION SCORE FOR EACH INDIVIDUAL AND EACH SEPARATE CRITERION SCORE AND THE SAME COEFFICENTS CORRECTED TO THE VARIATION OF THE MEAN SCORES OF THE TOTAL GROUP OF 183 STUDENTS

SUBJECT	NUMBER CASES	RAW CORRELATION COEFFICIENT	P. E.$_r$	CORRECTED CORRELATION COEFFICIENT
English	183	.777	.020	.777
Civics	26	.732	.061	.732
Algebra	144	.710	.028	.732
American History	180	.612	.031	.612
Physics	39	.800	.038	.747
French	30	.847	.036	.847
Solid Geometry	16	.412	.125	.403
Plane Geometry	128	.747	.027	.756
Trigonometry	22	.722	.073	.730
German	6	.750	.090	.728
Chemistry	27	.787	.053	.819
Latin	105	.709	.033	.737
Spanish	8	.675	.100	.858

All correlation coefficients between mean criterion scores and specific subject criterion scores (Table XXIII) are higher than comparable coefficients reported for IQ's and subject criterion scores (Table IX). This is to be expected because IQ's do not predict general levels of achievement perfectly.[5]

[5] Further light upon the fact that IQ-separate subject criterion correlation coefficients (Table IX) are smaller than mean criterion-separate subject criterion coefficients (Table XXIII) may be had by examining the profiles in Figures 14 and 15. The separate subject scores vary from the mean criterion score in both directions to the same degree. Accordingly, the algebraic difference between the mean and any specific subject score is relatively slight. However, the scores predicted by the IQ regression may all lie on one side of the actual subject scores, as in Figure 15, with resulting increased errors in the prediction.

In other words, the variation-corrected coefficients reported in Table XXIII may be looked upon as probable maximum values, for the data of this study, with the given degrees of reliabilities and dispersions, to be expected in the relationship between any index of general achievement and specific achievement.

IV. *The Efficiencies of Measures of General Achievement in Predicting Specific Achievement*

Variation-corrected correlation coefficients reported in the right-hand column of Table XXIII are expressed in the form of predictive efficiency indices in Table XXIV. The latter have been determined in a manner similar to that previously described for IQ's and criterion scores given in Table XVII. In no case does the predictive efficiency exceed 50 per cent. Achievements in ten of the thirteen subjects are predicted by mean criterion scores with efficiencies of less than 40 per cent.

TABLE XXIV

PER CENTS REDUCTION IN THE ERRORS OF PREDICTING SPECIFIC SUBJECT ACHIEVEMENTS USING MEAN CRITERION SCORES RATHER THAN CHANCE PLACEMENT.
(Hypothetical Subject Groups With the Same Variation of Mean Criterion Scores)

SUBJECT	PER CENT REDUCTION IN ERROR (E)	SUBJECT	PER CENT REDUCTION IN ERROR (E)
Civics	32.0	Algebra	32.0
English	37.1	German	31.5
Solid Geometry	8.5	Chemistry	42.7
French	46.7	Trigonometry	31.7
American History	21.5	Spanish	48.5
Plane Geometry	34.5	Latin	32.5
Physics	33.5		

The criterion scores are composed of average school marks for achievement estimated over periods extending from one semester to eight semesters, and to this average is added scores in standardized achievement tests with reliabilities approximating 90. (Table IV). Accordingly, the reliabilities[6] of the mean criterion scores are probably greater than that of IQ's in the

relatively narrow dispersion of IQ's in the upper twelfth grade. However, the differences in reliabilities of IQ's and mean criterion

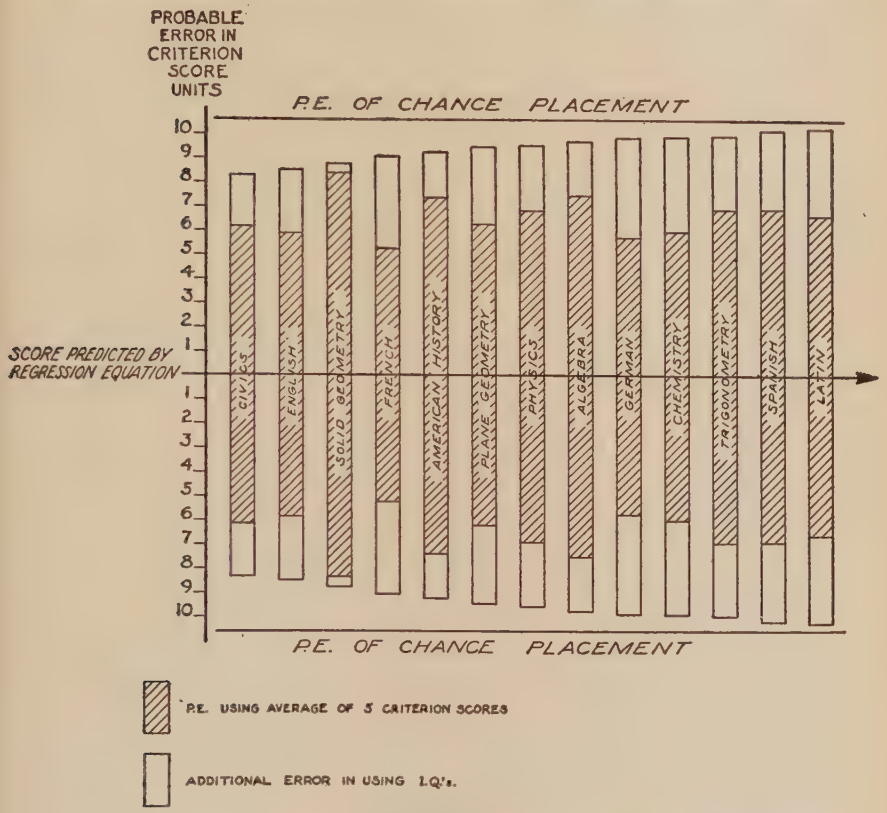


FIGURE 16

PROBABLE ERRORS OF THE ESTIMATE OF CRITERION SCORES USING: 1. THE AVERAGE OF FIVE CRITERION SCORES. 2. OTIS I.Q.

scores do not explain the greater predictive values of the latter.[7] The relatively low maximum efficiencies which can be reached

[6] From the Spearman-Brown prophecy formula (Holzinger, *op. cit.*, 169) and the estimated reliability of a single subject criterion score (p. 63), it can be shown that the minimum reliability of the mean of the 5 criterion scores is probably .92. However, the true reliability value of the average of the five criterion scores is likely to be greater than .92, probably approximating .95.

[7] See discussion of the effect of the reliability of the Otis Test on prediction, pp. 119–120.

by an index of general achievement in predicting specific achievement seems to place serious limitations upon their use. This limitation is depicted graphically in Figure 16. The IQ is used as such an index in Figure 16, which reports the probable error[8] of estimate for individuals of specific subject criterion scores through the use of the regression equation.[9] The error in estimate resulting from the use of IQ's is compared with (1) the error resulting from the use of the mean criterion score (assumed to be a perfect index of itself, average achievement) and (2) pure chance placement. The error of the IQ as a predictive index lies between these extremes in all instances given in Figure 16. In all subjects but solid geometry the error of the IQ as a predictive index approaches more closely the probable error of chance placement than it does the minimum error in predicting specific achievement scores by average achievement.

V. *The Ratio of the Variation of Achievement about Individual Levels to Variation about Group Levels of Achievement*

In addition to furnishing a means of determining the value of IQ's in predicting specific subject achievement, the data of Figure 16 also provide opportunity to estimate the extent of individual variation in specific achievement. The probable error of chance placement is a measure of the variation of the distribution of scores for different individuals in the same subject. The probable error of estimating the separate subject criterion scores by the mean criterion score is a measure of the variation of the scores in the separate subjects about their predicted achievement scores. Accordingly, there is afforded in Figure 16 a visual means of determining the ratio of the variation about predicted individual averages to variation about group averages.

Additional data are provided in Table XXIV for the same estimates. The per cent reductions in the variable errors of prediction of subject criterion scores (E) due to the use of the

---

[8] P. E. $_{est}$ = .6745 $\sigma\sqrt{1-r^2}$ where P. E. $_{est}$ = probable error of estimate of criterion scores, in a distribution whose standard deviation is σ; r is the correlation coefficient between the subject criterion scores and the predictive measure used, i. e., IQ or mean criterion score.

[9] See p. 100.

mean criterion score as a predictive index are given in Table XXIV. Derived data are reported in Table XXV as per cents of the total variation of the group attained by the variation of average individuals about the scores predicted by their own mean criterion scores. The process of derivation is to subtract the per cents of Table XXIV from 100 per cent.[10]

TABLE XXV

RATIOS OF THE STANDARD ERRORS OF ESTIMATING CRITERION SCORES FROM INDIVIDUAL MEAN CRITERION SCORES TO THE STANDARD DEVIATIONS OF THE COMPLETE DISTRIBUTIONS FOR 183 STUDENTS
(Expressed as Per Cents)

SUBJECT	PER CENT VARIATION	SUBJECT	PER CENT VARIATION
Solid Geometry	80.8	Civics	59.3
American History	71.5	Plane Geometry	58.9
Algebra	70.5	Chemistry	55.6
Physics	66.0	English	52.7
Trigonometry	64.2	German	52.3
Spanish	63.3	French	48.4
Latin	61.6		

Table XXV gives for each subject the per cent of the standard deviation for the complete distribution which equals the standard error of estimate, when the individual mean criterion scores are used as predictive devices. Similar per cents are reported in Table XXVI except that the variation of the individual criterion scores in the separate subjects is measured from the actual mean criterion scores. In Table XXV individual variation is measured from the criterion scores predicted by the individual means with the use of the regression equation. The per cents in Table

[10] The ratio of the partial sigmas to the total sigmas for the correlation coefficients reported in Table XXIII is also the ratio of the variation about predicted individual levels of achievement to variation about group levels of achievement. This ratio is $\frac{\sigma\sqrt{1-r^2}}{\sigma}$ in which σ is the standard deviation of the distribution of separate subject criterion scores and r is the correlation coefficient between mean criterion scores and the separate subject criterion scores.

$$\frac{\sigma\sqrt{1-r^2}}{\sigma} = \sqrt{1-r^2} = 1 - \left(1 - \sqrt{1-r^2}\right) = 1 - E$$

XXVI generally exceed those of Table XXV. This is because the individual variations given in Table XXV are measured with respect to the profile of the averages of the selected group of 183 students. The variation in individual profiles is reduced accordingly. Thus Table XXVI reports the actual variations from the individual means and may be looked upon as more representative of the true magnitudes of these variations.

TABLE XXVI

STANDARD DEVIATIONS OF THE SEPARATE SUBJECT CRITERION SCORES ABOUT INDIVIDUAL MEAN CRITERION SCORES, EXPRESSED AS PER CENTS OF THE STANDARD DEVIATIONS OF THE COMPLETE CRITERION SCORE DISTRIBUTIONS

SUBJECT	PER CENT VARIATION	SUBJECT	PER CENT VARIATION
Solid Geometry	113.0	Plane Geometry	66.9
Latin	91.3	Chemistry	63.7
Spanish	81.5	English	63.5
German	78.1	Algebra	59.5
Trigonometry	73.1	Physics	58.2
American History	73.0	French	55.3
Civics	69.9	ALL SUBJECTS	66.0[11]

The fact that all per cents in Table XXV, except that for French, and all per cents in Table XXVI including the total for all subjects exceed 50 may be interpreted as meaning that there is at least half as much variation in the separate subject scores of an average individual as there is in the scores for all individuals of these groups in any one subject. This fact is probably true even after corrections[12] for the reliabilities of the measures have been applied.

---

[11] This is the average (i. e., mean) of the standard deviations of 183 individual distributions of criterion scores, divided by the standard deviation of the distribution of the 183 criterion scores for the same individuals in English. This subject was chosen because it offered the only distribution in which all 183 individuals were represented. When the average of the individual standard deviations is divided by the standard deviation for average criterion scores, the per cent is 84.6.

[12] For example, the median corrected correlation coefficient in Table XXIII is .737 for Latin. Reliability coefficients of .80 and .95 for a single and the average criterion scores respectively may be assumed. The correction for attenuation will, accordingly, increase the "true" relationship for the median coefficient from .737 to .850. This corresponds to an individual variation which is 53 per cent of the group variation.

A comparison of the deviations of criterion scores from (1) individual and (2) group means is afforded by Figure 17. This diagram shows two distributions with the same 915 criterion scores in each. In the first distribution deviations are measured from all individual means. In the second distribution the same 915 criterion scores are given as deviations from the group means in each subject. Deviations from group means extended over a range of 16 intervals of 5 criterion score units each, 8 intervals

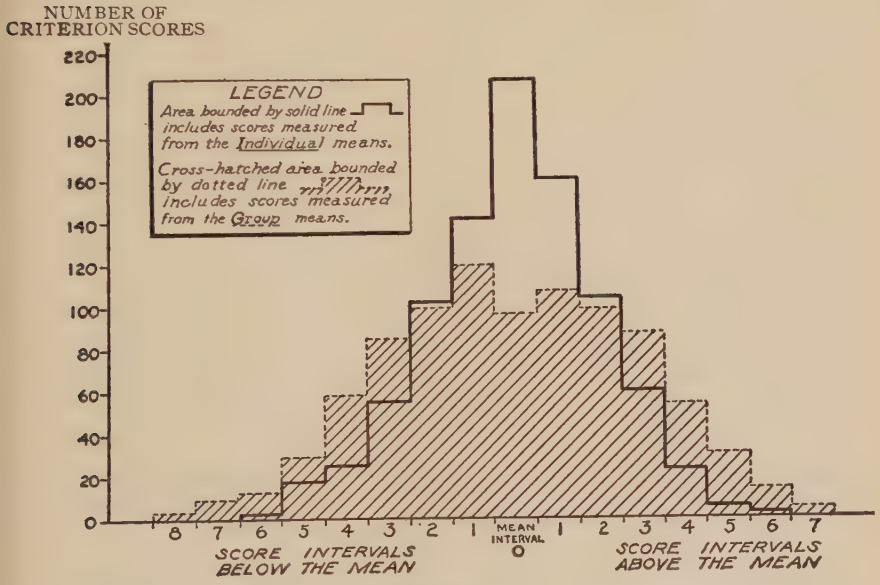


FIGURE 17

SUPERIMPOSED DISTRIBUTIONS OF 915 CRITERION SCORES FOR 183 INDIVIDUALS BASED UPON INTERVALS OF 5 CRITERION SCORE UNITS MEASURED FROM: 1. THE INDIVIDUAL MEANS 2. THE GROUP MEANS.

below the group mean interval and 7 intervals above the group mean interval. Deviations from individual means range over 13 intervals. Reading the diagram, two criterion scores were 6 intervals, or 30 criterion score units, below their individuals' mean criterion scores and 2 were 6 intervals above. Accordingly, individual deviations were extended over a range of 13 intervals, or 81.3 per cent of the entire range of deviations from group

means. This 81.3 per cent as the ratio of the entire range of variation about individual mean scores to the range of variation about the group mean scores may be compared with 66.0 per cent, the ratio of the corresponding standard deviations, reported in Table XXVI. The latter being the ratio of the standard deviations for average individuals to the standard deviation of the distribution of scores in a representative subject, English, may be considered as a more reliable measure. The standard deviation ratio is not subject to chance errors introduced by exceptional individual cases to the same extent as the ratio of the two entire ranges in Figure 17.

The magnitude of the variation of specific subject achievement is probably increased by unreliability of the measures. More accurate measuring techniques may be developed which may show less variation about mean levels of achievement than those reported in Tables XXV and XXVI. However, if the average individual's achievements in the restricted range of the academic subjects vary from their own means by an amount approximating and exceeding 50 per cent of the total variation of scores of individuals, it is evident that individual variation will be increasingly significant when non-academic subjects are included. Measures of achievement in physical education, music, shop work, art, home economics, typing, etc., were not included in this study. It is possible that their inclusion would further extend the range of variation in achievement about individual averages. The averaging of marks in these subjects with those in the different academic subjects, should be resorted to only with the recognition that this average has quite a restricted meaning. In a sense, it is as important in providing a measure of the total achievement of an individual to furnish estimates of the range of that individual's specific achievements as it is to furnish similar measures of the range of achievement for groups.

CHAPTER IX

Summary and Conclusions

Literature reporting the use of mental ability tests indicates that the basic uses of tests of mental ability in secondary schools and colleges have to do with the prediction of achievement. Some of these uses are concerned with: (1) educational and vocational guidances, (2) grouping pupils, (3) setting attainable individual standards of achievement, (4) evaluation of achievements of individuals or groups with respect to their probable capacities.

This study attempts to measure the degree of accuracy with which a single group test of mental ability, exemplified in the Otis Self-Administering Test of Mental Ability, Higher Examination, is capable of predicting academic achievements in specific subjects for average individuals.[1] Some of the purposes listed above, particularly, grouping and the evaluation of the achievements of groups, involve the prediction of the central tendencies of group performance. These are problems differing essentially from the prediction of the achievements of individuals. Mathematical analysis and practical experience indicate that the prediction of the central tendency of a group measure of achievement is subject to much less error than the prediction of the single measure for the achievement of an individual.

Mathematically, the smallest average error possible in predicting a number of different achievements from a single predictive index is the error involved in predicting these different achievements from their own average as the single index. Furthermore, the greatest possible average error for an infinite number of cases is that of prediction without an index, or chance placement. The average predictive error of the IQ lies between these extremes. This study attempts to determine within these limits the predictive value of an Otis IQ.

---

[1] Findings are obtained by the analysis of massed data, not by a series of case studies.

Data for the solution of this problem were furnished by the achievement records and mental ability test scores of 2449 students in the graduating classes of twelve senior high schools in Philadelphia. These records were collected in May 1928 as part of the cooperative research program of the Carnegie Foundation, the State Department of Public Instruction, and the Philadelphia Board of Public Education. The achievement record consisted of the average school mark for the entire time each subject was on the student's program in the senior high school. In addition to the average school mark, five standardized achievement tests were administered to each student in May 1928. Scores for these tests and the corresponding average school marks were combined to form composite subject criterion scores of the individuals in each subject group for whom the school records showed the same duration and recency of instruction. A modified T-score technique was used to scale the school marks and test scores which were converted into sigma scores. The arithmetic sum of these sigma scores is called the "criterion" score.

The Otis Self-Administering Test of Mental Ability Higher Examination, Form A, was administered with the achievement tests in May 1928. Scores in this test were used in the determination of IQ's by the Otis method. The fact that the IQ was determined at the same time as the criterion scores restricts the study to the making of comparison rather than prediction. However, studies of the constancy of the IQ indicate a probable correlation coefficient of .80 between IQ's determined at two intervals extending over periods as long as three years apart. Accordingly, the contemporary relationships determined in this study may be considered as probable maximum values for prediction of specific subject achievement over similar time intervals.

The solution of the problem was divided into seven steps as follows:

Question	Technique	Findings
1. What are the relationships between Otis IQ's and achievements in specific subjects?	1. (a) Pearson correlation coefficients were determined between Otis IQ's and criterion scores. (b) Correlation ratios (etas) were determined for five representative subject distributions. (c) Rectilinearity of relationship was proved by Blakeman's short test and by correspondence of r's and etas, for these five subjects. (d) Raw correlation coefficients were corrected for differences in the variation of IQ's for comparison with each other.	1. (a) Raw correlation coefficients ranged in 16 subjects from .575 for English to .123 for Spanish. The rank of subject groups in order of the magnitude of the correlation coefficients is as follows: (1) English, (2) social studies, (3) mathematics, (4) physical sciences, (5) foreign languages. (b) Corrected for differences in variation of IQ's the coefficients range from .575 for English to .142 for Spanish. (c) Comparison of the 16 raw correlation coefficients with 96 others reported by previous investigators shows that the correlation coefficients determined in this study rank in magnitude as follows: 8 above the middle, 2 at the middle, 6 below the middle of the respective distributions. (d) There is about as much variation in the findings of different investigators for correlations of measures of mental ability with achievement measures in any one subject as there is for all the different subjects in this investigation.
2. What is the value of an Otis IQ in the prediction of achievement in specific subjects?	2. Hull's predictive efficiency index, E, was determined for each subject criterion–IQ correlation coefficient. E is the per cent reduction in the variable error of prediction attained through the use of the IQ.	2. Predictive efficiencies range from 18.2 per cent for English to 0.8 per cent for Spanish. The reduction in the error of prediction is so small in amount that it fails to warrant the general use of a single Otis IQ as a sole predictive index of academic high school achievement in specific subjects.
3. What differences in predictive values exist for IQ's in different parts of	3. (a) The scatter diagrams for five specific subject criterion–IQ correlations were divided at the	3. (a) In algebra and physics the dispersion-corrected correlation coefficients were significantly

the IQ range? Is specific subject achievement predicted better by high or low IQ's?	interval nearest the mean IQ into an upper and lower IQ section. The lower IQ section included all IQ's below 110 in English, algebra, American history, and physics. Because the mean IQ was higher for the criterion group in Latin, IQ's below 116 constituted the lower IQ group in that subject. The upper IQ groups included all other IQ's in their respective distributions. Correlation coefficients were determined for these upper and lower IQ sections separately. The The coefficients were corrected for differences in variation of IQ's. (b) Regression line diagrams were prepared for five subjects and upon them were plotted the means of the arrays of subject criterion scores for each interval of 5 IQ units. The relationship of these means to the slope of the regression line of criterion scores upon IQ's was studied.	higher for upper compared with lower IQ groups. (b) In each case the mean criterion score for the highest IQ interval (130 and above) was higher than that predicted by the regression line. In three of five subjects, the mean criterion scores of the lowest IQ interval (below 85) were higher than scores predicted by the regression line. The reliability of this finding is substantiated in that these means are for score arrays which include 93 per cent of the cases. Accordingly, the IQ seems to be more valuable in predicting the achievement of the highest IQ groups than of the lowest IQ groups who persist in school through grade 12.
4. What is the extent of individual variation in subject achievement from individual average achievement?	4. Criterion scores were corrected to equivalent values in all subjects by raising or lowering the scales in conformity with mean IQ differences in the criterion score groups. All criterion scores were corrected to conform to the distribution of 2209 criterion scores in English through the use of the regression equation. A number of units was added to or subtracted from the raw criterion scores according to whether the given group mean IQ exceeded or was less than 107.3, the mean for the total group of 2209 students. The probable error of this technique approximated 2 score units for most subjects. 183 achievements profiles thus made available were studied.	4. Individuals vary greatly in the degree of unevenness of their achievement profiles. One individual had a Latin score in the lower 2 per cent of the distribution and an algebra score in the upper 16 per cent. Another profile shows almost uniform levels in all subjects. A correlation of $-.034 \pm .050$ between the means and sigmas of the 183 individual distributions indicates that there is no relationship between individual variation and average achievement.

5. How does the average standard deviation of individual scores compare with the standard deviation of the most complete group distribution?	5. The standard deviation of the 5 criterion scores in each profile was determined for 183 individuals. The average of these standard deviations was divided by the standard deviation of one complete distribution of scores in English for the same 183 individuals. This ratio was expressed as a per cent.	5. The profile of the average student showed 66 per cent as much variation about his mean criterion score as there was variation for 183 individuals about their group mean in the complete criterion score distribution for one subject. In no subject separately considered was individual variation less than half the group variation.
6. To what extent can achievements in specific subjects be predicted from individual average achievement?	6. Mean criterion scores were assumed to predict themselves (average achievement) perfectly. The correlation coefficients between mean criterion scores and each of the separate subject criterion scores were obtained and corrected for differences in variation of scores. Corrected coefficients were changed to predictive efficiency indices.	6. Specific subject achievement was predicted by average achievement with an efficiency of less than 40 per cent in 7 of 13 subjects. In 5 subjects the efficiency is between 40 and 50 per cent. In French, the predictive efficiency is 51.6 per cent.
7. How does the IQ compare with individual average achievement as an index for the prediction of specific subject achievement?	7. Predictive efficiency indices of IQ's and mean criterion scores were compared. In addition, probable errors of estimate were compared for both measures.	7. In all but one subject (solid geometry) the error of estimate in using the IQ as an index approximated chance placement. The improvement in predictive accuracy using average achievement rather than the IQ in the prediction of specific subject achievement is more than the improvement contributed by IQ's over chance placement.

I. *Critical Study of the Effect of Variation, Reliability, and Validity of Measures upon the Predictive Values of the IQ's*

The findings of any study are directly conditioned by the character of the data. Any study of prediction based upon the achievements of students in the twelfth grade must necessarily be one of the comparative success of students who have reached that grade. The average mental ability of those students who left school before the twelfth grade is probably lower than that of

those who continued. Evidence supporting this conjecture is furnished by a mean Otis IQ of 107.3 for the students tested in this program. Accordingly, with the elimination of some pupils, opportunity to judge of the success of the mental ability test data is restricted by the fact that this twelfth grade group probably displayed less heterogeneity of IQ's than groups in lower grades. It is in these lower grades that predictive devices are in demand. Lack of variation in the measures has an important effect upon correlation coefficients. The standard deviation of the Otis IQ's for the entire group studied is 9.6. It is quite likely that an unselected group of the same age range would exhibit a variation which may be expressed by a standard deviation of 12 IQ units.[2] The effect of this restriction in variation of IQ's may be estimated.[3] For example, in Table IX, p. 68, a correlation coefficient of .575 is reported between IQ's and English criterion scores. If the variation of IQ's were extended to that of a hypothetical unselected group and the variation of English criterion scores were similarly extended, and the same degree of relationship were maintained, the correlation coefficient corrected for increased dispersion would be .660. The predictive efficiency index similarly would be raised from 18.2 per cent to 24.9 per cent less error than chance placement. However, similarly extended variations of mean criterion scores would result in a similar increase in correlation coefficients between individual mean and specific subject criterion scores. This study has ranked the predictive efficiency of a single Otis IQ nearer the zero point of chance placement than the probable maximum efficiency attained by the individual achievement average. If IQ's and individual averages both increase the same degree in predictive efficiency, the ranking relationship will remain unchanged. Accordingly, the fact that the predictive errors in using IQ's more nearly approximated those of chance placement than they approximated the errors in predicting

---

[2] Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 97, *Report on Mental Ability of Philadelphia 12B pupils as measured by Scores in Otis Test of Mental Ability, Higher Examination*, p. 12.

[3] See pp. 69–70.

specific achievement with a measure of individual average achievement, probably is not due to the lack of variation of IQ's and achievement for the 12th grade group studied.

Another factor affecting relationships of data is the reliability of the measures. A minimum self-correlation of .712 and a probable value of .800 has been estimated for criterion scores in algebra and plane geometry.[4] These values are reasonably high for measures of school success which include school marks. If IQ's are used in predicting actual measures of school success, no corrections of IQ-criterion score correlation coefficients for lack of reliability of the criterion scores are desirable. However, it is entirely feasible to increase the reliability of the IQ determination by the administration of both forms of the Otis test. The resulting increase in reliability and consequently in predictive efficiency may be estimated. The probable error of a score in the Otis test is 2.56.[5] This indicates a reliability coefficient of .71 in the group under present consideration with a variation expressed by a standard deviation of 9.6 raw[6] score units. The administration of two forms of the test would raise the reliability coefficient to .83. Since Otis raw scores and IQ's are practically transmuted measures of the same quantity at the age level of the upper 12th grade, the reliability of the IQ determination similarly would be raised by the administration of two forms of the test. The correlation coefficients between IQ's and subject criterion scores correspondingly would be increased. For example, the increased reliability due to the use of both forms of the Otis test should result in an increase in the correlation coefficient between IQ's and English criterion scores estimated from the use of the Spearman correction for attenuation from .575 to .620. The predictive efficiency similarly would be raised from 18.2 to 21.5 per cent. The resulting increase in predictive efficiency is slight in comparison with 37.1 per cent found for

---

[4] See pp. 62-63.

[5] Otis Self-Administering Tests of Mental Ability, *Manual of Directions*, 12.

[6] Philadelphia Board of Public Education, Division of Educational Research and Results. Bulletin 97, *Report on Mental Ability of Philadelphia 12B pupils as measured by Scores in Otis Test of Mental Ability, Higher Examination*, 12. The standard deviation of IQ's is also expressed as 9.6 units. (See standard deviation for 2209 IQ's in Table X.)

the predictive value of the average criterion scores to predict achievement in English as reported in Table XXIV.

Not only does consideration of the reliability of the Otis test fail to explain the differences in the predictive powers of Otis IQ's contrasted with those of the average criterion scores, but also a study of the "validity" coefficients reported in Table V indicates that the validity of the Otis test for the current study is greater than the validity of the mental ability measures used in similar studies. The coefficient of .617 (with a restricted variation of IQ's indicated by a standard deviation of 9.0) between Otis IQ's and mean criterion scores is higher than any other correlation coefficient between Otis test results and criteria of school success found for other studies.

II. *Possibilities and Limitations of the use of Otis IQ's for Prediction of Specific Subject Achievement*

(a) *For Groups:* Examination of the scatter diagrams of the IQ's and subject criterion scores shows that, for IQ groups near the means of the IQ distributions (consequently those with the larger numbers of cases), mean subject criterion scores are predicted quite accurately by the IQ regression line. The prediction of mean achievement scores is desirable in grouping by ability, and in the evaluation of the achievement levels of groups. Accordingly, the use of an Otis IQ index is sanctioned for these purposes. Whether or not some index other than the IQ is more efficient is beyond the scope of this study.

Some slight reduction in the variation of achievement scores of groups may be effected by the use of the IQ as an index in grouping practice. This may be seen by examination of the criterion score arrays at successive IQ intervals in the scatter diagrams as in Table XXI. The average standard deviations of these arrays (partial sigmas) are less than the standard deviations of the complete distributions in every instance. However, in this respect, the relatively low correlation coefficients (and consequently high partial sigmas) between IQ's and subject criterion scores limit the value of the IQ index. Groupings as coarse as these indicated in Chapter VII, *i. e.*, two groups divided approximately at the mean IQ, show reductions in standard

deviations of criterion scores up to 13 per cent. Comparisons of the standard deviations for the upper and lower IQ groups in Table XX with those for similar total subject groups in Table X show these reductions. In all cases, except that of the upper IQ group of Latin students, criterion score variation of both the upper and lower group was less than the criterion score variation of the total group. However, the selection of relatively homogeneous distributions of criterion scores is not attained by IQ sectioning of the data in this study.

(b) *For Individuals*: The direct prediction of the specific subject achievements of individuals from a single Otis IQ is subject to very considerable error. In all specific subjects for which data were gathered in this study, the use of the IQ for prediction improved chance placement less than 20 per cent. The use of two forms of the Otis test should increase the predictive values to some slight degree, but not enough in most subjects to provide the IQ with predictive powers which approach those possible with a perfect index of average achievement. Accordingly, the use of the Otis IQ as a predictive index for the specific subject achievement of individuals seems limited to those cases where:

1. The only other possible prediction is equivalent to chance placement.

2. Very broad categories of achievement are used. As an illustration from the data of this study, no student with an IQ of 130 or over had a criterion score in any subject lower than one standard deviation below the group mean; accordingly, his probability of failure to be promoted in any subject was quite remote.

3. The range of achievement is wide and accurately measured permitting the establishment of these broad categories of achievement. For example, if achievement is measured over a range of several grades it is possible that higher predictive values than those indicated could be attributed to the IQ.

4. The IQ is properly weighted and combined with other predictive indices such as previous school marks, prognostic test scores, and the like, also properly weighted to form a composite predictive index; the most accurate method of determining weights involves the use of the multiple regression equation.

III. *The Significance of the Departure of Individual Subject Achievement from Individual Average Achievement.*

One of the significant findings of this study is that the criterion scores in the separate subjects of the individual profiles have an average standard deviation measured from the individual averages which is approximately two-thirds of the standard deviation of all criterion scores in a single subject (Table XXVI). In other words, the achievements of individuals are 66 per cent as variable about the individual average of achievement as they are about the group average of achievement. Even after corrections for selection and reliability have been made, it is probable that individual variation of scores will exceed one-half the amount of group variation. When it is remembered that these measurements have been restricted to the narrow range of the high school academic achievements of individuals, it is probable that the inclusion of comparable measurements of achievement in other subjects would further widen the variation of the achievements of individuals contrasted with those of groups. Such subjects are art, music, physical education, industrial arts, and home economics.

These facts seem to show that individual averages for all subjects combined into one measure have quite generally a restricted application, and for some students, almost no significance in indicating the probable degree of success in a given activity. Great care should be exercised in describing the total achievements of individuals. Such general terms as an "A student" or a "D student" may be quite misleading unless they are applied only to those individuals whose specific achievements vary slightly from this designation. This study gives little support to the use of such terms as a "low IQ" or a "high IQ" to imply low or high achievement in high school studies.

BIBLIOGRAPHY

Allen, W. S.: *A Study in Latin Prognosis.* Teachers College Contributions to Education No. 135. New York: Teachers College, Columbia University Bureau of Publications, 1923.

Allen combines a number of criteria using the multiple regression equation to predict achievement in Latin. The Latin prognosis is equally valid for English and mathematics.

Banker, H. J.: "Correlation Studies of the Students' Ability Index." *Journal of Educational Research* 20:31–37, June, 1929.

Develops a reliable substitute for mental ability tests using average school mark and chronological age.

Bates, G. M.: "The Work of the Student's Welfare Committee of Erasmus Hall High School." *Bulletin of High Points in the Work of the High Schools of New York City* 5:20–24, New York City Board of Education, June, 1923.

Description of the use of intelligence tests in the classification of pupils.

Billet, R. O.: "A Controlled Experiment to Determine the Advantage of Homogeneous Grouping." *Educational Research Bulletin* 7: 172, College of Education, Ohio State University, April 18, 1928.

Advocates use of I. Q. in grouping on basis of study in 9th grade in Harvey High School, Painesville, Ohio.

Book, W. F.: *Intelligence of High School Seniors.* New York: Macmillan Co., 1922.

A survey of the mental characteristics of high school seniors.

Branson, E. P.: "An Experiment in Arranging High School Sections on the Basis of General Ability." *Journal of Educational Research* 3: 53–55, January, 1921.

An attempt in differentiating the predictive power of high and low scores on the Otis scale for achievement in English.

Breed, F. S. and Breslich, E. R.: "Intelligence Tests and the Classification of Pupils II". *School Review* 30: 210–226, March, 1922.

Describes a reasonably successful but not permanent grouping on the basis of Otis test results.

Brown, A. W.: *The Unevenness of the Abilities of Dull and Bright Children.* Teachers College Contributions to Education No. 220. New York: Teachers College, Columbia University, Bureau of Publications, 1926.

Develops a T-score technique for comparing the achievement profiles of individuals on a variety of tests.

BUCKINGHAM, B. R.: "Mathematical Ability as Related to General Intelligence." *School Science and Mathematics* 21: 205–215, March 1921.

Reports and summarizes a number of correlation coefficients of intelligence test results and subject achievement at the college level.

BUREAU PUBLIC PERSONNEL ADMINISTRATION: "Measuring Abstract Intelligence in Personnel Work." *Public Personnel Studies* 7: 69–92. May-June, 1929.

Presents the inter-correlations among a number of group tests of intelligence. There is but little difference in the choice of group tests for personnel work.

BURR, M. Y.: *A Study of Homogeneous Grouping.* Teachers College Contributions to Education No. 457. New York: Teachers College, Columbia University, Bureau of Publications, 1931.

Shows composite profile charts for 100 pupils in 6th grade, divided into 3 groups according to educational age. Much overlapping is shown.

CATTELL, P.: "Comparability of I. Q.'s Obtained from Different Tests at Different I. Q. Levels." *School and Society* 31: 437–442, March 29, 1930.

Nine intelligence tests were compared. Little divergence for ratings in the middle of the scale, but there are significant divergences at the ends of the range.

——————— and GAUDET, F. J.: "The Inconstancy of the I. Q. as measured by Repeated Group Tests." *Journal of Educational Research* 21: 21–28, January, 1930.

Report at the end of the sixth year of Dearborn's Harvard study of the growth of school children. Gains on retests probably due to practice effects. Furnishes a description of Franzen's predictive indices.

CHAMBERS, G. G.: *An Introduction to Statistical Analysis.* New York: Crofts and Co., 1925.

Provides fundamental mathematical concepts for more advanced work in statistical methods.

CHAPMAN, J. C.: "The Unreliability of the Difference between Intelligence and Educational Ratings." *Journal of Educational Psychology* 14: 103–108, February, 1923.

Lack of predictive ability ascribed to unreliability.

CHEYNEY, W. W.: *The I. Q. and the Rate of Progress.* Philadelphia: University of Pennsylvania Ph. D. Thesis, 1930.

School success measured by the number of years required to complete the first eight grades correlated with I. Q.'s determined by group tests.

CLEM, O. M.: "Latin Prognosis." *Journal of Educational Psychology* 16: 160–169, March, 1925.

Reports more significant findings of his *Detailed Factors in Latin Prognosis*, Teachers College Contribution to Education, No. 144. A number of factors were correlated with success in Latin.

COLE, R. D.: "A Conversion Scale for Comparing Scores on 3 Secondary School Intelligence Tests." *Journal of Educational Research* 20: 190–199, October, 1929.

Provides transmutation tables for scores in Otis Self-Administering Higher Examination, Otis Group Advanced Scale, and Terman Group Test.

COLLEGE ENTRANCE EXAMINATION BOARD: *Annual Reports.* New York: 1926 and 1928.

Provides statistical analysis of Scholastic Aptitude Test including correlation coefficients between results in the test and marks in college.

COLVIN, S. S.: "Intelligence Tests and Their Use." *21st Yearbook of the National Society for the Study of Education.* Bloomington, Ill.: Public School Publishing Company, 1922.

Most complete summary up to the date of publication.

CONDIT, P. M.: "The Prediction of Scholastic Success by Means of Classification Examinations." *Journal of Education Research* 19: 331–335, May, 1929.

Comparison of scores on the Thurstone Psychological Examination and average scholastic ratings of students at Colorado State Teachers College.

COX, P. W. L.: "Providing for Individual Differences by Means of Grouping by Ability." *Ninth Annual Schoolman's Week Proceedings.* Philadelphia: University of Pennsylvania Bulletin 23: 233–244, 1922.

Reports practice and results of grouping pupils in junior and senior high schools.

CRAFTS, L. L.: "Causes of Failure in Plane Geometry as Related to Mental Ability." *Mathematical Training* 16: 481–492, December, 1923.

Attributes failure to lack of mental ability as measured by the Terman test.

CRAWFORD, A. B. (Director—Department of Personnel Study—Yale): "Forecasting Freshman Achievement." *School and Society* 31: 126–132, January 25, 1930.

Shows school record to be more successful than College Entrance Examination Board averages and scores in the Scholastic Aptitude Test. Reliability of .80 reported for freshman grades at Yale.

DEARBORN, W. F.: *Intelligence Tests.* New York: Houghton-Mifflin Co., 1928.

A significant recent text.

DeVoss, J. C.: "Specialization of the Abilities of Gifted Children." Ch. 12 in Terman, L. M.: *Genetic Studies in Genius* 1: 307–362, Stanford University, California: Stanford University Press, 1925.

Applies method developed by Kelley to the Stanford Achievement scores of 100 gifted children and finds a similar result, *i. e.*, 29 per cent of the differences between scores are in excess of chance elements due to unreliability of the tests. Gifted children have slightly but not significantly more specialization of "abilities" than normal children.

Dickson, V. E.: *Intelligence Tests and School Reorganization.* Yonkers, New York: World Book Co., 1922.

A pioneer study of one of the uses of intelligence tests.

———— *Mental Tests and the Classroom Teacher.* Yonkers, New York: World Book Co., 1923.

Significant for this study in summarizing data for the constancy of the I. Q.

———— "The Use of Mental Tests in School." *Administration Monographs.* 4. Berkeley, California, Board of Education, June, 1922.

Advocates the use of the I. Q. as a fairly reliable prognostic measure.

Dunlap, K.: "Mental Tests." *Progressive Education* 7: 57–67, March, 1930.

The usefulness of an intelligence test varies directly with the definiteness with which the performances measured are related to the purpose of the test.

Dvorak, A.: "Recognition of Individual Differences in the Junior High School." *School Review* 30: 679–685, November, 1922.

A questionnaire study reporting results from 86 schools in different parts of United States. Significant in reporting the numbers of schools using different bases for ability grouping.

Flemming, C. W.: *A Detailed Analysis of Achievement in the High School; Comparative Significance of Certain Mental, Physical and Character Traits for Success.* Teachers College Contributions to Education No. 196. New York: Teachers College, Columbia University, Bureau of Publications, 1925.

Reports intercorrelations for school marks, teachers' estimates of intelligence, school attitude and energy, chronological age, Terman and Otis tests. Obtains a multiple correlation coefficient of .8454 in estimating school marks.

Freeman, F. N.: *Mental Tests.* New York: Houghton-Mifflin Co., 1926.

An important text.

Fuller, Violet: *The Relation of Intelligence to the Passing of Secondary School Subjects.* Boulder, Colorado: University of Colorado M. A. Thesis, 1924.

Reports correlation coefficients for 800 students at State Preparatory School, Boulder, Col., between I. Q.'s and school marks in 23 courses. The coefficients vary from .05 for algebra to .62 for general science.

GARRETT, H. E.: *Statistics in Psychology and Education.* New York: Longmans, Green and Company, 1926.

A simplified text in measurement.

GATES, A. I.: "The Correlations of Achievement in School Subjects with Intelligence Tests and Other Variables." *Journal of Educational Psychology* 13: 129–139; 223–235; 277–285, March, April, May, 1922.

Correlations of achievement and intelligence scores for elementary school subjects in grades 1–8 at Scarborough School, New York.

GATES, A. I. and LA SALLE, J.: "The Relative Predictive Values of Certain Intelligence Tests Together with a Study of Effect of Educational Achievement Upon Intelligence Test Scores." *Journal of Educational Psychology* 14: 517–539, December, 1923.

Reports correlation coefficients reduced from .868 to .755 after 0 and 20 months respectively between National Intelligence Test Scores and the average of 4 scholastic tests.

HENMON, A. C. and OTHERS: *Prognosis Tests in the Modern Foreign Languages.* Modern Foreign Language Study. New York: Macmillan Co., 1929.

Reports results of a number of investigations in foreign language prognosis.

HOLZINGER, K. G.: *Statistical Method in Education.* Philadelphia: Ginn and Co., 1928.

A reasonably complete text of statistical method.

———— *Statistical Tables for Students in Education and Psychology.* Chicago: University of Chicago Press, 1925.

Reports tables of squares, products, quotients, logarithms, and statistical constants.

HUFFAKER, C. L.: "Predictive Significance of the Correlation Coefficient." *Journal of Educational Research* 21: 46–48, January, 1930.

Develops formula for Hull's E, predictive efficiency index, corrected for unreliability in the criterion.

HUGHES, J. M.: "The Use of Tests in the Evaluation of Factors which Condition the Achievement of Pupils in High School Physics." *Journal of Educational Psychology* 16: 217–231, April, 1925.

Used Otis Test of Mental Ability, Higher Examination, for 29 schools. Compares distribution of Otis I. Q.'s with a normal distribution of I. Q.'s.

HULL, C. L.: *Aptitude Testing.* Yonkers, New York: World Book Co., 1928.

A most complete statistical treatment of prognostic testing.

JORDAN, A. M.: "Correlations of Four Intelligence Tests with Grades." *Journal of Educational Psychology* 13: 419–429, October, 1922.

Correlation coefficients for Otis, Army Alpha, Miller, and Terman tests with high school subject marks.

———— "The Validation of Intelligence Tests." *Journal of Educational Psychology* 14: 348–366; 414–428, September and October, 1923.

Reports averages of 50 correlation coefficients for four intelligence tests and average school marks. These vary from .38 for Army Alpha to .56 for the Miller.

JORDAN, J. N.: "Prognosis in Foreign Languages in Secondary Schools." *School Review* 33: 541–546, 1925.

Compares correlation coefficients between teachers' marks in Latin, French and Spanish and each of 3 predictive scores: Terman, National, and prognosis tests.

KAULFERS, W. V.: "The Intelligence Factor in Foreign Language Achievement." *School Review* 39: 42–48, January, 1931.

Reports for 1002 junior high school students in San Diego and Los Angeles the r's between I. Q.'s (Terman both forms) and grade-points averages in Spanish for 6 different semesters.

KEFAUVER, G. N.: "The Validity of Bases for Forming Ability Groups." *Teachers College Record* 31: 99–114, November, 1929.

Author was responsible for determining the bases for distributing pupils to ability groups in 7th grade of the junior high schools of Fresno, California. Teachers' estimates of capacity show the highest relationships with general success.

KELLEY, T. L.: "A New Method for Determining the Significance of Difference in Intelligence and Achievement Scores." *Journal of Educational Psychology* 14: 321–333, September, 1923.

A significant contribution in the determination of the reliability of unevenness in the Stanford achievement profiles of 96 eighth grade children.

———— *Interpretation of Educational Measurements.* Yonkers, N. Y.: World Book Co., 1927.

Provides a theoretical and practical means of selecting an educational test for any specific purpose.

———— *Statistical Method.* Yonkers, N. Y.: World Book Co., 1923.

A fundamental mathematical reference and textbook in statistical method.

KORNHAUSER, A. W.: "Test and High School Records as Indicators of Success in any Undergraduate School of Business." *Journal of Educational Research* 16: 342–356, December, 1924.

Reports validity coefficients for Otis Test of Mental Ability Higher Examination.

LAWS, A. R. and BOWIE, S.: "Intelligence Tests in Examinations for Junior Scholarships." *Journal of Experimental Pedagogy and Training College Record* 6: 155–169, December 5, 1921.

Reports correlation coefficients between intelligence test results and terminal marks.

MAY, M. A.: "Predicting Academic Success." *Journal of Educational Psychology* 14: 429–440, October, 1923.

Correlation of intelligence test scores with "honor points."

MILLER, W. S.: "General Intelligence Tests." *School Review* 28: 94, February, 1920.

Describes classification of entering students in University of Minnesota high school according to intelligence test results with a list of favorable results.

———— "The Administrative Use of Intelligence Tests in the High School." *National Society for the Study of Education 21st Yearbook* 189–222, Bloomington, Ill.: Public School Publishing Co., 1922.

A study of the validity of the Miller Mental Ability Test.

MINER, J. R.: *Tables of* $\sqrt{1-r^2}$ *and* $1-r^2$ *for Use in Partial Correlation and in Trigonometry*, Baltimore, Md.: Johns Hopkins Press, 1922.

Four-place tables as described.

ODELL, C. W.: "An Annotated Bibliography Dealing with the Classification and Instruction of Pupils to Provide for Individual Differences." *University of Illinois Bulletin* 21, November 19, 1923.

Furnishes a rather complete bibliography to the date of publication. Enough notations are given to facilitate selection.

———— *Educational Statistics.* New York: Century Co., 1925.

A standard text.

———— *Predicting the Scholastic Success of College Freshmen.* University of Illinois Bulletin 37, Bureau of Educational Research, 1927.

Reports correlation coefficients between I.Q.'s and success in foreign languages.

————*The Use of Intelligence Tests as a Basis of School Organization and Instruction.* University of Illinois Bulletin 20, Bureau of Educational Research, December 25, 1922.

Reports correlation coefficients between achievement and intelligence.

———— *Educational Measurement in High School.* New York: Century Co., 1930.

The most recent textbook in the field of secondary school measurement.

———— *A Selected Annotated Bibliography Dealing with Examinations*

and School Marks. University of Illinois Bulletin 24, Bureau of Educational Research, January 15, 1929.

A guide to the literature.

———— *Predicting the Scholastic Success of College Students.* University of Illinois Bulletin 52, Bureau of Educational Research, September 30, 1930.

Offers hope for multiple correlation coefficients of .75 or .80 with specific subject grades in college and .80 or over with average grades.

OMANS, A. C.: "Provision for Ability Grouping in Junior and Senior High School." *American School Board Journal* 65: 55–58, October, 1922.

Questionnaire from 50 schools in different parts of United States. Grouping by ability more favored than individual promotion.

ORLEANS, J. S. and SOLOMON M.: *Latin Prognosis Test*—Manual of Directions. Yonkers, N. Y.: World Book Co., 1928.

Correlation coefficients of .80 with a criterion of Latin ability have been obtained.

OTIS, A. S.: *Statistical Methods in Educational Measurements.* Yonkers, N. Y.: World Book Co., 1925.

A valuable text. Presents a most complete discussion of the use of percentiles.

———— *Otis Self-Administering Test*—Manual of Directions. Yonkers, N. Y.: World Book Co., 1922.

Reports validity, reliability, and utility of Otis tests.

PHILADELPHIA BOARD OF PUBLIC EDUCATION—DIVISION OF EDUCATIONAL RESEARCH AND RESULTS: *Mental Ability of 12B Students.* Bulletin 97, 1928.

Gives data for Otis Test results used in this study.

————*Subject-Matter Tests—Cooperative Program.* Bulletin 109, 1928.

Gives school summaries of achievement test data used in this study.

————*Senior High School Subject Promotions.* Bulletin 194, January, 1931.

Cites data showing bases of school marks and reasons for variation.

————*Junior High School Subject Promotions.* Bulletin 195, January, 1931.

Cites data for bases of school marks.

———— *American Council Alpha Tests and Final Marks in French* (4), *German* (4), *and Spanish* (4). Bulletin 201, January, 1931.

Reports data for the reliability of school marks.

POWERS, S. R.: "Intelligence as a Factor in the Selection of High School Subjects." *School Review* 30: 452–455, June, 1922.

Shows brighter students elect mathematics, Latin, natural science,

Spanish and French, manual training and mechanical drawing, history, commercial and domestic art, in that order.

PRESSEY, S. L.: "The Efficiency of a Group Scale of Intelligence in Prognosticating Success and Failure in the Junior High School." *Journal of Applied Psychology* 3: 381–385, December, 1919.

Describes use of Indiana Group Point Scale in Bloomington. Results are rather highly correlated with failure and success in school work.

PROCTOR, W. M.: "Psychological Tests as a Means of Measuring the Probable School Success of High School Pupils." *Journal of Educational Research* 6: 258–270, October, 1920.

States that 50 per cent of below normal pupils will be eliminated in 2 years.

———— *Psychological Tests and Guidance of High School Pupils. Journal of Educational Research Monographs No. 1.* Bloomington, Ill.: Public School Publishing Co., 1923.

Reports correlation coefficients for first year high school marks with Binet I. Q. and Army Alpha.

———— "The Use of Intelligence Tests in the Educational Guidance of High School Pupils." *School and Society* 8: 473–478, 502–509, October 19 and 26, 1918.

Concludes that the Stanford Revision apparently affords as good a measure of success in school as do teachers' estimates or previous school marks. Pupils with I. Q.'s of 95 or below are not likely to be successful in high school work.

———— "The Use of Psychological Tests in the Educational Guidance of High School Pupils." *Journal of Educational Research* 2: 533–546, September, 1920.

Shows that pupils guided in the choice of subjects according to Stanford Revision results show less elimination and failure than do those who are not so guided.

PYLE, W. H.: "The I. Q. and the Individual." *School and Society* 26: 788–790, December 17, 1927.

Finds considerable scatter in learning excellence in the same individuals.

RECTOR, W. G.: "A Study in the Prediction of High School Success." *Journal of Educational Psychology* 16: 28–38, January, 1925.

Reports correlation coefficients between Army Alpha I.Q.'s and marks in 15 specific subjects for 431 high school students.

REEVE, W. D.: "Homogeneous Grouping of High School Students by Means of Psychological Tests." *4th Yearbook of the National Association of Secondary School Principals:* 81–94, Menasha, Wis.: George Banta Publishing Co., 1921.

Discusses correlation of tests and school marks.

ROGERS, AGNES L.: "Mental Tests as a Means of Selecting and Classifying College Students." *Journal of Educational Psychology* 11: 181–192, April, 1920.

Scores for 300 Goucher students on Thorndike and Rogers Tests were correlated with school marks. Coefficients were low for the single tests but high when the test results were combined.

ROSS, C. C.: *The Relation Between Grade School Record and High School Achievement.* Teachers College Contributions to Education No. 166. New York: Teachers College, Columbia University, Bureau of Publications, 1925.

Correlations between Terman group test and school success during the first year in high school.

———— and HOOKS, N. T.: "How Shall We Predict High School Achievement?" *Journal of Educational Research* 22: 184–195, October, 1930.

Summarizes a number of studies showing that the previous academic record is superior to teachers' estimates of intelligence, aptitude tests and general intelligence test results.

RUCH, G. M. and STODDARD, G. D.: *Tests and Measurements in High School Instruction.* Yonkers, N. Y.: World Book Co., 1927.

Standard text reporting critical statistical data for most secondary school tests existing prior to its publication.

RUGG, H. O.: *Statistical Methods Applied to Education.* New York: Houghton-Mifflin Co., 1917.

A pioneer text in the field of education measurement.

RUGG, L. S.: "Retests and the Constancy of the I. Q." *Journal of Educational Psychology* 16: 381–393, May, 1925.

States that clearly the obtainable data point to the conclusion that the I. Q. remains relatively constant.

RUNNELS, R. C.: *Manual for Determining the Equivalence of Mental Ages Obtained from Group Intelligence Tests.* Test Method Helps, Department of Psychology and Test Service. Yonkers, N. Y.: World Book Co., 1930.

Gives data for the transmutation of National Intelligence Test, Scale A, Form 1; Haggerty Intelligence Examination-Delta 2; Otis Group Intelligence Scale, Advanced A; Terman Group Test of Mental Ability, Form A; Otis Self-Administering Test of Mental Ability, Higher Examination, Form A.

SHIDELER, J. W.: "A Correlation of Teachers' Grades and the Scores of Intelligence Tests." *School Review* 29: 733–734, December, 1921.

Terman group test given to 170 high school pupils at Fort Scott, Kansas, and correlated with marks in the various subjects. Concludes that in-

telligence test results should be used to supplement the teacher's judgment. Concludes that intelligence test results should be used to supplement the teacher's judgment.

SKAGGS, E. B.: "The Correlation of General Intelligence Tests and School Standing." *Journal of Educational Psychology* 2: 169–171, March, 1920.

1530 cases in a state normal school gave a correlation coefficient of .31.

SPEARMAN, C.: "The Proof and Measurement of Association between Two Things." *American Journal of Psychology* 15: 72–101, January, 1904.

Description of correlation and the correction for attenuation.

SNOW, W. C.: *A Study of the Relationship between Mental Status and Classroom Achievement of Public School Pupils.* Chicago: University of Chicago, Master's Thesis, 1925.

Reports the correlation coefficient between Illinois Intelligence scores for September and May.

SYMONDS, P. M.: "The Accuracy of Certain Standard Tests for School Sectioning and Marking." *Journal of Educational Psychology*, October, 1924.

Statistical discussion of the error in dichotomic (two group) sectioning.

———— *Measurement in Secondary Education.* New York: Macmillan Co., 1927.

A practical text on standardized tests and their use in the high school.

———— *Ability Standards for Standardized Achievement Tests in the High School.* New York: Teachers College, Columbia University, Bureau of Publications, 1927.

Provides tables of standards in nine achievement tests for students of given mental ability scores in the Terman group test.

————"Homogeneous Grouping." *Teachers College Record* 32: 501–517, March, 1931.

Summarizes the evidence for and the recent evidence against homogeneous grouping including the evidence for specificity of achievement. "On the average a twelve-year-old child in school will show a discrepancy of 1.8 years or grades between two academic subjects (1.5 years with true measures), or if a child is taking five subjects the chances are even that he will show as much as 4.1 years or grades between some two of his subjects. (3.6 years with true measures.)"

THORNDIKE, E. L.: *An Introduction to the Theory of Mental and Social Measurements.* New York: Teachers College, Columbia University, Bureau of Publications, 1914.

———— "Intelligence and its Uses." *Harper's Magazine* 140: 227–235, 1919.

An article describing intelligence tests for laymen.

TILDSEY, J. L.: "Some Possibilities Arising from the Use of Intelligence Tests." *Bulletin of High Points in the Work of the High Schools of New York City* 3: 4-10, June, 1921.

Tells of uses of intelligence tests in New York City.

TOOPS, H. A.: "The Use of Intelligence Tests by Universities." *School and Society* 24: 603, July 17, 1926.

Questionnaire to 110 colleges and universities throughout the United States.

———— "The Status of University Intelligence Tests in 1923–24." *Journal of Educational Psychology* 17: 23–36; 110–124, January-February, 1926.

Reports frequency of correlation coefficients between intelligence scores and freshman scholarship marks.

WALLIN, J. E. W.: "Intelligence Irregularity as Measured by Scattering on the Binet Scale." *Journal of Educational Psychology* 13: 140–151, March, 1922.

Finds scatter greater for normal than for sub-normal children.

WASHBURNE, C. W.: *Adapting the Schools to Individual Differences.* Part II, 24th Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1925.

Finds that ability grouping is not sufficient to care for the irregularities in the subject achievements of individuals.

WINTER, O.: "Chicago Intelligence Test in Harrison Technical High School." *School Review* 28: 772–775, December, 1920.

It was found that the correlation between intelligence test scores and certain achievement tests and also school marks was rather low.

WOODY, CLIFFORD: "Measurement of the Effectiveness of Differentiation of High School Pupils on the Basis of the Army Intelligence Tests." *Journal of Educational Research* 7: 397–409, May, 1923.

A study of the achievement of freshmen in a small Michigan high school shows that if the class had been divided into 3 groups according to scores in Army Alpha Test the division would have agreed fairly well with achievement as measured by tests in English, algebra, and Latin.

APPENDIX A

Program for a Study of the Relations of Secondary and Higher Education. Abstract of a Report to the Joint Commission of the Association of College Presidents and of the State Department of Public Instruction in the State of Pennsylvania. Carnegie Foundation for the Advancement of Teaching, 522 Fifth Avenue, New York, 1928, pp. 6–9.

Procedure to be followed in the Examination of Twelfth-grade Pupils

(1) *The Tests.* The tests proposed for the study of the twelfth-grade pupils are of the so-called "new type." Each consists of a large number of questions covering a particular field in a graded sequence of difficulty and requiring answers on the "true-false," or "completion," or "multiple choice" principle. Most of those that will be used are new tests prepared for the occasion. They are easily understood, simple of administration, mechanically and quickly scored. They test chiefly exact knowledge and the power of accurate reasoning. Guessing is heavily penalized, so that when the tests are properly administered the results are believed to be more truly indicative of the qualities tested than in the usual type of examination with subjective ratings. The power of organization and of orderly exposition, and certain other "higher processes" are not directly measured. Experience with the tests has increasingly shown, however, that the "higher processes" are directly conditioned by the comprehensiveness of important knowledge, and by the clearness with which the ideas or elements in such processes are held. It is believed that these tests will furnish additional light on such relationships. There is, of course, no other type of test that lends itself so well to standard measurement from school to school.

(2) *Pupils to be Tested.* It is desired to make a fully representative study of members of the final class (equivalent to the twelfth grade) in high schools and other secondary schools of Pennsylvania. To accomplish this it will be necessary to reach every individual so classified regardless of the curriculum pursued (technical, commercial, college preparatory, etc.) and regardless also of his immediate prospects of graduation.

(3) *Selection of Subjects.* It is expected that in the cooperating schools each pupil will take at least five tests—one in each of the following groups: English, mathematics, foreign languages, social sciences, and natural sciences, although pupils who have never studied algebra or a foreign language may compensate with other examinations in the social or natural sciences.

The examination in English and in American history will be expected of every pupil regardless of what English or history courses he may have had or when he took them, if at all.

Each student who has ever studied mathematics other than arithmetic will take the algebra examination; if he has had both algebra and plane geometry, he will take the algebra and the plane geometry tests, unless he has also taken solid geometry or trigonometry or both, in which case he will take the test in solid geometry or in trigonometry in place of the test in plane geometry. Pupils will thus divide into four groups according as they take: (1) algebra; (2) algebra and plane geometry; (3) algebra and solid geometry; and (4) algebra and trigonometry.

In foreign languages a pupil may choose between Latin, French, German, Spanish, and Italian. He may take tests in more than one if he so desires. If he has never studied a foreign language he may take an additional test in the social sciences or natural sciences.

In social sciences American history will be expected of all pupils, as explained above. If he wishes, a pupil may take also one of the following: ancient history, general European history, elementary economics, or civics and American government.

In the natural sciences each student who has studied biology, or physics, or chemistry during the last three years of his course will be expected to take the examination in one of these three subjects. If he has studied no one of these subjects he will take the test in general science, regardless of what other instruction in science he may have had, unless permitted by his principal to substitute an additional test in foreign languages or in the social sciences.

(4) *Arrangement of the Program.* It will devolve upon the principal, or his deputy, in each co-operating school to determine what is to be the examination program of each member of his senior class.

The hours fixed for the various examinations and the amount of time allowed for each subject appear in the schedule below. Within this program the principal is at liberty to make such adjustments as he sees fit, provided that all the tests in a given group are taken within the limit of the half-day session under which they are listed. If it be thought desirable in the interests of the student morale to give the tests later to absentees this may be done in case the papers of such pupils are plainly marked *absentee* and returned in a separate package. The days and hours fixed by the schedule are the same as those appointed for the college tests described later. It is considered desirable by superintendents and principals who have been consulted that the schedule be rigidly adhered to as to day and session grouping. Nevertheless, in the few cases where it is desired that a pupil take more than two tests that are scheduled for the same session, it is expected that the principal will give the third test

at some convenient later period, after taking due precautions that all test-questions regularly distributed have been called in. The exact hours of each session may be adjusted to suit the local program of the school concerned.

Schedule of Examinations for Senior Pupils in Secondary Schools

		Hours
MONDAY, MAY 7		
9 a. m.	*English*	2 (required of all pupils)
2 p. m. (a)	*Foreign Languages*	
	(1) French	1½
	(2) German	1½
	(3) Italian	1½
	(4) Latin	1½
	(5) Spanish	1½
(b)	*Natural Sciences*	
	(1) General Science	1
	(2) Biology	1
	(3) Physics	1½
	(4) Chemistry	1½
TUESDAY, MAY 8		
9 a. m.	*Mathematics*	
	(1) Algebra	1¾
	(2) Plane Geometry	1
	(3) Solid Geometry	1
	(4) Trigonometry	1
2 p. m.	*Social Sciences*	
	(1) American History	1½ (required of all pupils)
	(2) Ancient History	1½
	(3) General European History	1½
	(4) Elementary Economics	1½
	(5) Civics and American Government	1½

(6) *Administration of the Tests.* Each school desiring to collaborate in the enterprise here described will report through its superintendent, or if a private school, through its principal, to the State Department of Public Instruction, on blanks to be furnished for the purpose, the number of copies of each test that will be required. This order must be filed with the State Department on or before March 24 to ensure its being filled. Orders received later will be filled if possible.

In all cases where the orders are so filed the Carnegie Foundation will undertake to pay one-half of the costs. This will enable the school to secure the tests at a price of not more than three cents each, or an average of less than twenty cents per pupil. In cases where schools are unable to purchase tests even at this low figure, the State Department will undertake to make arrangements whereby such schools can secure the tests free of charge. This, however, will be done only when it is absolutely necessary in order to ensure state-wide participation in the project.

Each principal will receive a manual of directions and a scoring key for each kind of test that he orders. The tests should be scored by the teachers under the immediate supervision of the principal or of his responsible representative. The school should make a record of the scores in each test for local purposes, and the test papers should then be forwarded at once to the State Department of Public Instruction at Harrisburg.

Both principal and teachers should bear in mind that the results from all the tests will be considered as strictly confidential. They will be studied and analyzed as a whole or in large groups and no school or teacher will be exposed to any comparisons nor will the data be reported in such a way that interested parties could make comparisons of an individual sort between teachers or groups of teachers.

It should be stated likewise that every effort will be made to do justice to the individual pupil. The time and extent of his study of each subject will be recorded and due allowance made for the interval of "forgetting" as well as for certain other factors that might affect his scores.

(7) *Cooperation of the Pupil.* It is obvious that the success of the proposed project depends wholly on the extent to which each participating school supports it not only with its authority but also with every possible effort to secure a serious and *bona fide* result. Any other outcome would vitiate the returns far more than non-participation.

Naturally the tests must be made a requirement of the school. They afford such ample provision for the lower ranges of ability that inexcusably poor results, especially when due to a lack of serious effort, would justify the school in questioning promotion or graduation. The validity of the scoring of the answers is completely in the hands of the school authorities; it is not a question, therefore, of an "outside" examination.

To the great majority of students appeal can be made on quite different grounds. They are asked to participate in a first-class intellectual adventure involving at the same moment all of the high schools and other secondary schools as well as the colleges of the State. Interscholastic games they know, but the opportunity for a school to match its intellectual achievement with that of others on even terms has hitherto been lacking. Without too much concern

for courses taken or forgotten here is a contest that pits the young student at the close of his high school experience, against the measure of a statewide performance and points the moral that what counts is neither "credits" nor "units" but intelligence based upon an abundance of clear and available ideas.

Notable success in a test of this character would assuredly imply distinction of a high order, and schools could well give local recognition for a fine achievement either in separate subjects or in total scores. It can hardly be doubted that most schools would heartily welcome this opportunity of emphasizing with their students the practical significance of that phase of their activities which, though usually central, is often obscured by more spectacular interests.

APPENDIX B

An additional source of data was a "Pupil's Information Blank" returned by the schools for each student tested. A copy of this blank is reproduced on p. 141 in order to indicate the source of some of the data used:

THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

IN COOPERATION WITH THE JOINT COMMISSION OF
THE ASSOCIATION OF PENNSYLVANIA COLLEGE PRESIDENTS
AND THE STATE DEPARTMENT OF PUBLIC INSTRUCTION

Study of Relations of Secondary and Higher Education in Pennsylvania

PUPIL'S INFORMATION SHEET

SECTION I. Directions to the Principal:

1. All three sections of this blank should be carefully filled out before May 5th, partly by the School authorities but chiefly by each member of your senior class—12th Grade—the same pupils who will take the tests on May 7th and 8th. Please see that *ink* is used.
2. After the blanks have been filled out by the pupil, the School should enter in the column headed "Average School Mark" an estimate of the character of the work done by the pupil in each subject that he lists as having been taken in school. This estimate should be one of the following letters: A, for good; B, for average; C, for poor; or F, for failure, and should cover with a single letter all the work done in a single subject.
3. After being filled out by the pupil, Sections II and III should be detached and retained at the School. After the tests have been given and scored, Section I, duly filled out by both pupil and School, should be placed with the assembled test papers of the pupil whose name it bears and should be forwarded with them to the Carnegie Foundation as directed.

Name.....Age on May 1, 1923.....Sex.....
(Print last name) (First name) (Middle name) (Years) (Months) (Boy or Girl)

School. . . . . City or Town.....County.....

Name of Parent, (if living) or Guardian. .....Exact address.....
(Street and Number)

City, Town or County.....State.....

Name and address of one other person, either relative or friend, who would be likely to know where you are in case you change your present address next year:

Name .....Exact Address .....
(Street and Number)

City or Town. . . . . State .....

Nationality of parents (that is, are they *by birth* English, or German, or American, or Italian, etc.?) Father?..... Mother?.....

What was, or is, your father's occupation? .....Did he attend high school?.....College?.....

Did your mother attend high school? College? .....Did brothers or sisters attend college?.....Graduate?.....

Do you intend to go to college? ..If so, where? .....When?.....

If you do not intend to go to college, is it because (1) you lack funds? . . (2) family needs your support?.... (3) you are not interested?.....

Have you secured a position? What sort of work do you prefer to do? .....Do you prefer working to studying?.....

What school subjects have you liked best? .....What do you like best to do outside of school?.....

In the diagram below place one or more X's to the left of each subject that you have studied in high school or secondary school so as to show in which of the eight semesters you studied it.

Column I, below, to be filled by the Principal. Columns II and III should be left blank.

1st year		2nd year		3rd year		4th year			Average School Mark	Test Scores					Ratings		
1st Sem.	2nd Sem.	1st Sem.	2nd Sem.	1st Sem.	2nd Sem.	1st Sem.	2nd Sem.			1	2	3	4	Total	I	II	III
								English									
								French									
								German									
								Italian									
								Latin									
								Spanish									
								General Science									
								Biology									
								Physics									
								Chemistry									
								Algebra									
								Plane Geometry									
								Solid Geometry									
								Trigonometry									
								Amer. History									
								Ancient History									
								Gen. European History									
								Element. Economics									
								Civics and Government									

(The Pupil will not write on the next five lines)

Occupation:
First Year—X
Second Year—O

1. Student in Penna.
2. Student not in Penna.
3. Clerk in Office
4. Stenographer and Typist
5. Merchandising
6. Industry and Manufacturing
7. Transportation
8. Civil Service
9. Farm Worker
10. Home Worker
11. Married
12. Miscellaneous

Student in Pennsylvania. Institution.....

Student not in Pennsylvania. Institution.....

APPENDIX C

Analysis of content in Otis Self-Administering Test of Mental Ability Higher Examination, Form A.

I. English—Total Number of Items..... 28
(a) Vocabulary not included in other categories..... 19
1. The opposite of hate is (?)
1. enemy, 2. fear, 3. love, 4. friend, 5. joy
4. The opposite of honor is (?)
1. glory, 2. disgrace, 3. cowardice, 4. fear, 5. defeat
6. Quiet is related to sound in the same way that darkness is related to (?)
1. a cellar, 2. sunlight, 3. noise, 4. stillness, 5. loud
9. The opposite of economical is (?)
1. cheap, 2. stingy, 3. extravagant, 4. value, 5. rich
16. A meal always involves (?)
1. a table, 2. dishes, 3. hunger, 4. food, 5. water
17. Of the five words below, four are alike in a certain way. Which is the one not like these four?
1. bend, 2. shave, 3. chop, 4. whittle, 5. shear
18. The opposite of never is (?)
1. often, 2. sometimes, 3. occasionally, 4. always, 5. frequently
23. A contest always has (?)
1. an umpire, 2. opponents, 3. spectators, 4. applause, 5. victory
27. The opposite of awkward is (?)
1. strong, 2. pretty, 3. short, 4. graceful, 5. swift.
32. If the settlement of a difference is made by mutual concession, it is called a (?)
1. promise, 2. compromise, 3. injunction, 4. coercion, 5. restoration
34. Of the five things below, four are alike in a certain way. Which is the one not like these four (?)
1. smuggle, 2. steal, 3. bribe, 4. cheat, 5. sell
36. The opposite of hope is (?)
1. faith, 2. misery, 3. sorrow, 4. despair, 5. hate
39. What people say about a person constitutes his (?)
1. character, 2. gossip, 3. reputation, 4. disposition, 5. personality
47. What is related to few as ordinary is to exceptional (?)
1. none, 2. some, 3. many, 4. less, 5. more
48. The opposite of treacherous is (?)

1. friendly, 2. brave, 3. wise, 4. cowardly, 5. loyal

49. Which one of the five words below is most unlike the other four?
1. good, 2. large, 3. red, 4. walk, 5. thick

57. Which of the following is a trait of character?
1. personality, 2. esteem, 3. love, 4. generosity, 5. health

71. A man who is influenced in making a decision by preconceived opinions is said to be (?)
1. influential, 2. prejudiced, 3. hypocritical, 4. decisive, 5. impartial

74. A statement the meaning of which is not definite is said to be (?)
1. erroneous, 2. doubtful, 3. ambiguous, 4. distorted, 5. hypothetical

(b) Reading Comprehension (non-incidental)..... 9

11. Which one of the six statements below tells the meaning of the following proverb? "The early bird catches the worm."
1. Don't do the impossible.
2. Weeping is bad for the eyes.
3. Don't worry over troubles before they come.
4. Early birds like worms best.
5. Prompt persons often secure advantages over tardy ones.
6. It is foolish to fret about things we can't help.

12. Which statement above tells the meaning of this proverb? "Don't cry over spilt milk."

13. Which statement above explains this proverb?
"Don't cross a bridge till you get to it."

29. Which one of the six statements below tells the meaning of the following proverb? "The burnt child dreads the fire."
1. Frivolity flourishes when authority is absent.
2. Unhappy experiences teach us to be careful.
3. A thing must be tried before we know its value.
4. A meal is judged by the dessert.
5. Small animals never play in the presence of large ones.
6. Children suffer more from heat than grown people.

30. Which statement above explains this proverb? "When the cat is away, the mice will play."

31. Which statement above explains this proverb? "The proof of the pudding is in the eating."

41. If the words below were arranged to make a good sentence, with what letter would the second word of the sentence begin? Make it like a printed capital.
same means big large the as

43. Suppose the first and second letters in the word CONSTITUTIONAL were interchanged, also the third and fourth letters, the

fifth and sixth, etc. Print the letter that would then be the twelfth letter counting to the right.

70. If the words below were arranged to make the best sentence, with what letter would the last word of the sentence end? Print the letter as a capital.
sincerity traits courtesy character of desirable and are

II. Mathematics—Total Number of Items..... 19

(a) Arithmetic Problems..... 10

2. If 3 pencils cost 5 cents, how many pencils can be bought for 50 cents?
7. A party consisted of a man and his wife, his two sons and their wives, and four children in each son's family. How many were there in the party?
15. If a boy can run at the rate of 6 feet in ¼ of a second, how far can he run in 10 seconds?
35. If 10 boxes full of apples weigh 400 pounds, and each box when empty weighs 4 pounds, how much do all the apples weigh?
40. If 2½ yards of cloth cost 30 cents, what will 10 yards cost?
45. If 4½ yards of cloth cost 90 cents, what will 2½ yards cost?
56. If a strip of cloth 24 inches long will shrink to 22 inches when washed, how long will a 36-inch strip be after shrinking?
61. If Frank can ride a bicycle 30 feet while George runs 20 feet, how far can Frank ride while George runs 30 feet?
72. A hotel serves a mixture of 2 parts cream and 3 parts milk. How many pints of cream will it take to make 15 pints of the mixture?
75. If a wire 20 inches long is to be cut so that one piece is 2-3 as long as the other piece, how long must the shorter piece be?

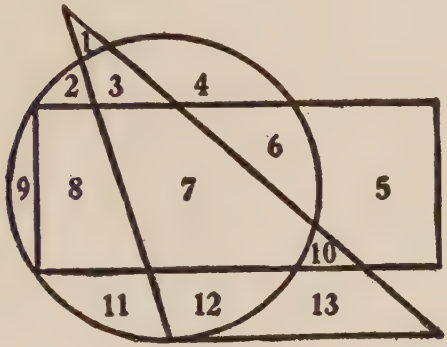
(b) Number Concepts..... 4

21. One number is wrong in the following series.
What should that number be
1 4 2 5 3 6 4 7 5 9 6 9
24. Which number in this series appears a second time nearest the beginning?
6 4 5 3 7 8 0 9 5 9 8 8 6 5 4 7 3 0 8 9 1
44. One number is wrong in the following series.
What should that number be?
0 1 3 6 10 15 21 28 34
60. One number is wrong in the following series.
What should that number be?
1 3 9 27 81 108

(c) Line and Space Concepts..... 5

54. What is related to a cube in the same way in which a circle is related to a square?
    1. circumference, 2. sphere, 3. corners, 4. solid, 5. thickness

65. What number is in the space which is in the rectangle and in the triangle but not in the circle?



66. What number is in the same geometrical figure or figures as the number 8?

67. How many spaces are there that are in any two but only two geometrical figures?

68. A surface is related to a line as a line is to (?)
    1. solid, 2. plane, 3. curve, 4. point, 5. string

III. Natural Science—Total Number of Items..... 9

(a) Biology..... 4

3. A bird does not always have (?)
    1. wings, 2. eyes, 3. feathers, 4. a nest, 5. a bill

5. A fox most resembles a (?)
    1. wolf, 2. goat, 3. pig, 4. tiger, 5. cat

8. A tree always has (?)
    1. leaves, 2. fruit, 3. buds, 4. roots, 5. a shadow

73. What is related to blood as physics is to motion?
    1. temperature, 2. veins, 3. body, 4. physiology, 5. geography

(b) Physics..... 1

19. A clock is related to time as a thermometer is to (?)
    1. a watch, 2. warm, 3. a bulb, 4. mercury, 5. temperature

(c) Astronomy..... 2

25. The moon is related to the earth as the earth is to (?)
1. Mars, 2. the sun, 3. clouds, 4. stars, 5. the universe

52. The statement that the moon is made of green cheese is (?)
1. absurd, 2. misleading, 3. improbable, 4. unfair, 5. wicked

(d) Hygiene..... 1

33. What is related to disease as carefulness is to accident?
1. doctor, 2. surgery, 3. medicine, 4. hospital, 5. sanitation.

(e) Chemistry..... 1

53. Of the five things following, four are alike in a certain way. Which is the one not like these four? 1. tar, 2. snow, 3. soot, 4. ebony, 5. coal

IV. Social Science—Total Number of Items..... 9

(a) Economics..... 1

10. Silver is more costly than iron because it is (?)
1. heavier, 2. scarcer, 3. whiter, 4. harder, 5. prettier

(b) History..... 1

14. An electric light is related to a candle as an automobile is to (?)
1. a carriage, 2. electricity, 3. a tire, 4. speed, 5. glow

(c) Logic..... 4

22. If the first two statements following are true, the third is (?) All members of this club are Republicans. Smith is not a Republican. Smith is a member of this club.
1. true, 2. false, 3. not certain

42. If the first two statements following are true, the third is (?) George is older than Frank. James is older than George. Frank is younger than James.
1. true, 2. false, 3. not certain

50. If the first two statements following are true, the third is (?) Some of Brown's friends are Baptists. Some of Brown's friends are dentists. Some of Brown's friends are Baptist dentists.
1. true, 2. false, 3. not certain

69. If the first two statements following are true, the third is (?) One cannot become a good violinist without much practice. Charles practices much on the violin. Charles will become a good violinist.
1. true, 2. false, 3. not certain

(d) Social Relations..... 3

46. A man's influence in a community should depend upon his (?)
1. wealth, 2. dignity, 3. wisdom, 4. ambition, 5. political power

59. Revolution is related to evolution as flying is to (?)
1. birds, 2. whirling. 3. walking, 4. wings, 5, standing

63. A man who is averse to change and progress is said to be (?)
1. democratic, 2. radical, 3. conservative, 4. anarchistic, 5. liberal

V. Foreign Languages—Total Number of Items 0

VI. Not Classified—Total Number of Items 10

20. Which word makes the truest sentence?
Men are (?) shorter than their wives.
1. always, 2. usually, 3. much, 4. rarely, 5. never

26. Which word makes the truest sentence?
Fathers are (?) wiser than their sons.
1. always, 2. usually, 3. much, 4. rarely, 5. never

28. A mother is always (?) than her daughter.
1. wiser, 2. taller, 3. stouter, 4. older, 5. more wrinkled

37. If all the odd-numbered letters in the alphabet were crossed out, what would be the tenth letter not crossed out? Print it. DO NOT MARK THE ALPHABET.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

38. What letter in the word SUPERFLUOUS is the same number in the word (counting from the beginning) as it is in the alphabet? Print it.

51. How many of the following words can be made from the letters in the word LARGEST, using any letter any number of times?
great, stagger, grasses, trestle, struggle, rattle, garage, strangle

55. If the following words were seen on a wall by looking in a mirror on an opposite wall, which word would appear exactly the same as if seen directly?
1. Ohio, 2. saw, 3. noon, 4. motor, 5. otto

58. Find the two letters in the word DOING which have just as many letters between them in the word as in the alphabet. Print the one of these letters that comes first in the alphabet.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

62. Count each N in this series that is followed by an O next to it if the O is not followed by a T next to it. Tell how many N's you count.

N O N T Q M N O T M O N O O N Q M N N O Q N O T O N A M O N O M

65. Print the letter which is the fourth letter to the left of the letter which is midway between O and S in the alphabet?

Grand Total 75

WESTBROOK PUBLISHING CO.
WPC
PHILA.